

# Bildungsverläufe und Stichprobenselektivität: Analysen zur Stichprobenselektivität des Mikrozensuspanels 1996-1999 am Beispiel bildungsstatistischer Fragestellungen

Schimpl-Neimanns, Bernhard

Veröffentlichungsversion / Published Version

Dissertation / phd thesis

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

## Empfohlene Zitierung / Suggested Citation:

Schimpl-Neimanns, B. (2008). *Bildungsverläufe und Stichprobenselektivität: Analysen zur Stichprobenselektivität des Mikrozensuspanels 1996-1999 am Beispiel bildungsstatistischer Fragestellungen*. (GESIS-Forschungsberichte - Reihe Sozialwissenschaftliche Datenanalyse, 1). Bonn: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-262236>

## Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

## Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## GESIS-Forschungsberichte

Reihe Sozialwissenschaftliche Datenanalyse

Band 1

Bernhard Schimpl-Neimanns

### **Bildungsverläufe und Stichprobenselektivität**

Analysen zur Stichprobenselektivität des Mikrozensuspanels  
1996 -1999 am Beispiel bildungsstatistischer Fragestellungen



## **Bildungsverläufe und Stichprobenselektivität**

GESIS-Forschungsberichte

Reihe Sozialwissenschaftliche Datenanalyse

Band 1

Herausgegeben von GESIS

GESIS ist ein Institut der Leibniz-Gemeinschaft.

**Bernhard Schimpl-Neimanns**

## **Bildungsverläufe und Stichprobenselektivität**

Analysen zur Stichprobenselektivität des Mikrozensuspanels  
1996 -1999 am Beispiel bildungsstatistischer Fragestellungen

GESIS-Forschungsberichte

Reihe Sozialwissenschaftliche Datenanalyse, Band 1

GESIS 2008

## **Bibliographische Information Der Deutschen Bibliothek**

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Die vorliegende Veröffentlichung ist im Februar 2008 unter dem gleichen Titel vom Fachbereich Sozial- und Kulturwissenschaften der Justus-Liebig-Universität Gießen als Dissertation angenommen worden. Die Arbeit entstand im Rahmen des 2003 bis 2005 durchgeführten Projektes „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“, dem der Verfasser angehörte. Die Partner im Verbundprojekt waren das Statistische Bundesamt, das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen und die Freie Universität Berlin. Das dieser Arbeit zugrunde liegende Teilprojekt des Zentrums für Umfragen, Methoden und Analysen (ZUMA) wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 07SWF06B gefördert.

ISSN 1867-2876

ISBN 978-3-86819-003-8

Herausgeber, Druck und Vertrieb:  
GESIS e.V.

Lennéstraße 30, 53113 Bonn

Tel.: 02 28 / 22 81 - 0

[info@gesis.org](mailto:info@gesis.org)

Printed in Germany

© 2008 GESIS. Alle Rechte vorbehalten. Insbesondere ist die Überführung in maschinenlesbare Form sowie das Speichern in Informationssystemen, auch auszugsweise, nur mit schriftlicher Einwilligung des Herausgebers gestattet.

## Zusammenfassung

Der Mikrozensus ist als rotierende Panelstichprobe angelegt, bei der die Haushalte eines Auswahlbezirkes vier Jahre lang befragt werden. Aus den Querschnittserhebungen des Mikrozensus kann daher auch ein Panel generiert werden. Das Mikrozensuspanel eröffnet viele neue inhaltliche und methodische Analysemöglichkeiten und kann dazu beitragen, das Defizit an Längsschnittdaten zu verringern.

Da nach dem Prinzip der Flächenstichprobe wegziehende Personen und Haushalte nicht weiter befragt, sondern durch die nachziehenden Personen bzw. Haushalte ersetzt werden, ist damit potenziell eine Selektivität der Stichprobe verbunden. In Anbetracht des dringenden Bedarfes längsschnittlicher Informationen in der empirischen Bildungsforschung konzentriert sich diese Arbeit auf bildungsstatistische Zusammenhänge für die überdurchschnittlich räumlich mobile Teilpopulation Jugendlicher und junger Erwachsener.

Die Selektivitätsfragen werden am Beispiel von drei Übergängen untersucht: Dem Auszug aus dem Elternhaus, dem Besuch der gymnasialen Oberstufe bis zum Abitur und der beruflichen Ausbildung im dualen System bis zum Abschluss und Übergang in den ersten Beruf.

Für die Ausfallanalysen werden Selektionsmodelle für kategoriale Daten verwendet. Diese Modelle erlauben nicht nur die Untersuchung, in welcher Weise die inhaltlich interessierenden Analysevariablen zusammenhängen, sondern auch, ob Ausfälle und Statuswechsel voneinander abhängen. Ausfallanalysen sind in hohem Maße von der Gültigkeit der bei der Modellierung getroffenen Annahmen abhängig. Zur Validierung dieser Annahmen ist deshalb die Verwendung externer Referenzdaten unerlässlich. Für Bildungsverläufe der gymnasialen Oberstufenschüler bildet die amtliche Bildungsstatistik die Referenzstatistik. Für das Thema Bildungsqualifikation und Berufseinstieg wird die IAB-Regionalstichprobe 1975-2001 verwendet.

Insofern bei Panels auf die Angaben aus vorherigen Befragungen zurückgegriffen werden kann, stellt das Mikrozensuspanel auch wichtige Informationen zur Qualitätsbeurteilung bereit. Die Ergebnisse zur Konsistenz der Bildungsangaben im Mikrozensus belegen, dass im Mikrozensus sowohl die Zahl der Besucher der gymnasialen Oberstufe als auch die Zahl der Abiturienten im Vergleich mit der Bildungsstatistik mit erheblichen Klassifikationsfehlern behaftet sind.

Bei den Ausfallanalysen zeigt sich, dass die Stichprobenselektivität bei Bildungsverläufen zum Besuch und Abschluss der gymnasialen Oberstufe dem Typ bedingt zufälliger Ausfälle entspricht. Sie ist weitgehend mithilfe von Gewichtungsfaktoren korrigierbar, wenn Zusammenhänge zwischen der beruflichen Stellung der Familienbezugsperson der Jugendlichen und dem Ausfall berücksichtigt werden. Während die Stichprobenselektivität somit korrigierbar ist, bleiben die ermittelten Abiturquoten aber aufgrund der Klassifikationsfehler verzerrt.

Die Annahme bedingt zufälliger Ausfälle trifft allerdings für berufliche Ausbildungsverläufe nicht zu. Erfolgreiche Absolventen haben ein geringeres Ausfall- bzw. Umzugsrisiko als Ausbildungsabbrecher. Der Ausfall hängt somit vom interessierenden Übergang des Ausbildungsstatus ab und entspricht dem Typ nicht ignorierbarer Ausfälle. Mit dem Selektionsmodell können unter Berücksichtigung dieser Zusammenhänge dennoch wirksame Ausfallkorrekturen geschätzt werden. Jedoch scheitern Gewichtungsfaktoren, die auf der (Standard-) Annahme bedingt zufälliger Ausfälle basieren. Diese führen zu inhaltlich von der Referenzstatistik abweichenden Ergebnissen.

Vor dem Hintergrund, dass nachträgliche Korrekturverfahren immer mit Einschränkungen verbunden sind, wird die Frage aufgeworfen, wie das Potenzial des Mikrozensus für die Längs-



schnittnutzung verbessert werden könnte. Um die durch das gegenwärtige Stichprobendesign entstehenden Schwierigkeiten zu reduzieren oder ganz zu vermeiden, wäre unter anderem eine Weiterbefragung räumlich mobiler Personen anzustreben.

# Inhalt

Zusammenfassung . . . . .	5
Vorwort . . . . .	9
1 Einleitung . . . . .	11
1.1 Aufgabenstellung der Arbeit . . . . .	12
1.2 Konzeption und Aufbau . . . . .	13
2 Datenbeschreibung . . . . .	17
2.1 Mikrozensuspanel 1996-1999 . . . . .	17
2.1.1 Stichprobenplan und Erhebungsprogramm . . . . .	17
2.1.2 Zusammenführung der Querschnittsangaben . . . . .	19
2.2 Amtliche Bildungsstatistik . . . . .	22
2.3 IAB-Regionalstichprobe 1975-2001 . . . . .	22
3 Methodische Grundlagen . . . . .	25
3.1 Klassifikation der Ausfallprozesse . . . . .	25
3.3 Selektionsmodelle für kategoriale Daten . . . . .	27
3.3 Beispiele . . . . .	29
3.3.1 Vollständig zufälliger Ausfall (MCAR) . . . . .	31
3.3.2 Bedingt zufälliger Ausfall (MAR) . . . . .	32
3.3.3 Nicht ignorierbarer Ausfall (NINR) . . . . .	33
3.3.4 Modelle mit externen Referenzdaten . . . . .	36
4 Auszug aus dem Elternhaus . . . . .	43
4.1 Theoretische Konzepte und Forschungsstand . . . . .	43
4.2 Abgrenzung der Stichprobe und Variablenkonstruktion . . . . .	47
4.2.1 Abgrenzung der Stichprobe . . . . .	47
4.2.2 Variablenkonstruktion . . . . .	50
4.3 Verlaufsdatenanalysen . . . . .	54
4.3.1 Das diskrete Hazardraten-Regressionsmodell . . . . .	54
4.3.2 Modellergebnisse . . . . .	57
4.3.3 Diskussion der Ergebnisse . . . . .	60
4.4 Zusammenfassung . . . . .	69
5 Vom Besuch der gymnasialen Oberstufe bis zum Abitur . . . . .	71
5.1 Ausgangsinformationen und Vorgehensweise bei der Ermittlung des Abitur- abschlusses . . . . .	72
5.2 Deskriptive Ergebnisse zum Besuch der gymnasialen Oberstufe und dem Abschluss mit dem Abitur . . . . .	74

5.3	Zur Datenqualität der Bildungsangaben im Mikrozensus im Vergleich zur Bildungsstatistik . . . . .	83
5.3.1	Methode und Abweichungsgründe . . . . .	83
5.3.2	Absolventen mit allgemeiner Hochschulreife und Fachhochschulreife in der Bildungsstatistik und im Mikrozensuspanel . . . . .	85
5.3.3	Vergleiche zum Besuch der gymnasialen Oberstufe im Mikrozensus und in der Bildungsstatistik . . . . .	87
5.3.4	Antwortstabilität der Angaben zum allgemein bildenden Schulabschluss . . . . .	92
5.4	Analysen zum Ausfall der Besucher der gymnasialen Oberstufe bis zur Ermittlung des Abiturs . . . . .	94
5.4.1	Ansätze zur Erklärung sozial ungleicher Bildungsbeteiligung . . . . .	94
5.4.2	Statistische Modelle zu dem bis 1999 erreichten Abschluss . . . . .	95
5.4.3	Statistische Modelle für einzelne Statusübergänge . . . . .	101
5.5	Zusammenfassung . . . . .	108
6	Berufliche Ausbildungsverläufe bis zum Übergang ins Erwerbsleben . . . . .	111
6.1	Abgrenzung der Auszubildenden und der Ausbildungsabschlüsse . . . . .	112
6.1.1	Ausgangsinformationen im Mikrozensuspanel . . . . .	113
6.1.2	Ausgangsinformationen in der Beschäftigtenstichprobe . . . . .	114
6.1.3	Konstruktion vergleichbarer Abgrenzungen . . . . .	114
6.2	Auszubildende und Ausbildungsabschlüsse im Mikrozensuspanel und in der Beschäftigtenstichprobe . . . . .	120
6.2.1	Hypothesen zum Ausfall und deskriptive Analysen zum Übergang 1996/97 . . . . .	122
6.2.2	Modellergebnisse zum Ausfall im Mikrozensuspanel . . . . .	128
6.3	Können Gewichte die Stichprobenselektivität korrigieren? . . . . .	135
6.4	Chancen der Beschäftigung im erlernten Beruf . . . . .	138
6.4.1	Theoretische Konzepte und Forschungsstand im Überblick . . . . .	139
6.4.2	Abgrenzung der Stichprobe und Variablenkonstruktion . . . . .	140
6.4.3	Deskriptive Analysen zum Berufswechsel . . . . .	143
6.5	Zusammenfassung . . . . .	146
7	Zusammenfassung und Ausblick . . . . .	149
7.1	Stichprobenselektivität bei Bildungsverlaufsanalysen . . . . .	150
7.2	Ausblick . . . . .	156
	Literaturverzeichnis . . . . .	161
	Anhang . . . . .	177
	Tabellenverzeichnis . . . . .	201
	Abbildungsverzeichnis . . . . .	205

# Vorwort

Die vorliegende Untersuchung basiert zu großen Teilen auf überarbeiteten und ergänzten Arbeiten, die im Rahmen meiner Tätigkeit im Verbundprojekt "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe" entstanden sind. Das Projekt wurde durch eine aufeinander abgestimmte Förderung des Forschungsverbundes seitens des Bundesministeriums für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) ermöglicht.

Das dieser Arbeit zugrunde liegende Teilprojekt des Zentrum für Umfragen, Methoden und Analysen (ZUMA) wurde von 2003 bis 2005 mit Mitteln des BMBF unter dem Förderkennzeichen 07SWF06B gefördert. Die Partner im Verbundprojekt waren das Statistische Bundesamt (Mikrozensusgruppe, Bonn, und Gruppe Mathematisch-statistische Methoden, Wiesbaden), das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (Gruppe Interfachliche Erhebungen und Analysen) und die Freie Universität Berlin (Institut für Statistik und Ökonometrie, Prof. Dr. U. Rendtel).

Mein Dank gilt dem BMBF für die finanzielle Förderung des Projekts sowie ZUMA für die Möglichkeit, das Projekt durchzuführen und im Anschluss daran die Ergebnisse für diese Arbeit aufzuarbeiten.

Im Verbundprojekt, bei ZUMA und darüber hinaus habe ich von vielen inhaltlichen und methodischen Anregungen profitiert. Mein Dank gilt den folgenden Personen: Edin Basic, Stefan Bender, Michael Blohm, Michael Braun, Siegfried Gabler, Robert Herter-Eschweiler, Kristina John, Paul Lüttinger, Rolf Porst, Ulrich Pötter, Julia Schroedter, Stefan Weick, Heike Wirth und Christof Wolf. Für die technische Umsetzung des Textes und die Erstellung der Druckvorlage danke ich Bettina Zacharias.

Nicht zuletzt danke ich ganz herzlich den Gutachtern Prof. Dr. Peter Schmidt und Prof. Dr. Ulrich Rendtel für konstruktive Kommentare zu früheren Fassungen der Arbeit.



# 1 Einleitung

In der sozialwissenschaftlichen Forschung weisen viele Fragestellungen einen längsschnittlichen bzw. zeitlichen Bezug auf. Beispielsweise fragt die Arbeitsmarkt- und Armutsforschung nach den Risiken der Arbeitslosigkeit und der Verbleibsdauer in prekären Lebenslagen. Die Sozialstrukturforschung beschäftigt sich mit beruflichen und sozialen Auf- und Abstiegen und ihren Veränderungen im Zeitverlauf. Für die empirische Bearbeitung solcher Fragen sind Querschnittdaten, die nur Angaben zu einem Zeitpunkt liefern, nur insofern verwendbar als sie gesellschaftliche Entwicklungstrends abbilden können. Um darüber hinaus auch Aussagen über individuelle Veränderungen, wie z. B. über die Dynamik der häuslichen Arbeitsteilung im Eheverlauf (siehe Schulz und Blossfeld 2006) treffen zu können, werden längsschnittliche Informationen der gleichen Untersuchungseinheiten benötigt. Aus der Perspektive der gesellschaftlichen Dauerbeobachtung und Sozialberichterstattung liegen die Forschungspotenziale von Längsschnittdaten darin, gesellschaftliche Entwicklungen sowie Risiken und Chancen sich wandelnder Lebenslagen genauer zu untersuchen (Schupp et al. 1996: 42). Weil in individuellen Längsschnittdaten Angaben zur abhängigen und zu unabhängigen Variablen vor und nach einem Ereignis vorliegen, besteht die Chance, nicht beobachtete Merkmale statistisch zu kontrollieren. In dieser Hinsicht sind individuelle Längsschnittdaten auch für kausale Analysen und für die Evaluation von sozialpolitischen Programmen einsetzbar. Zumeist werden Längsschnittdaten in Form von Paneluntersuchungen erhoben, bei denen zu bestimmten Zeitpunkten dieselben Fragen denselben Untersuchungseinheiten gestellt werden. Paneldaten liefern i. d. R. zeitdiskrete Verlaufsangaben (vgl. zur Typisierung von Längsschnittdaten und deren Analysepotenziale u. a. Blossfeld et al. 1986; Blossfeld und Prein 1998; Martin et al. 2006; Singer und Willett 2003.)

Für Deutschland wurde von der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI 2001: 255, 296) und im ersten Armuts- und Reichtumsbericht der Bundesregierung (BMA 2001: 37) ein Defizit an Längsschnittdaten festgestellt. Ähnlich wie bei den Paneldaten des U.S.-amerikanischen Current Population Survey (CPS; siehe Madrian und Lefgren 2000) oder des Labour Force Survey (LFS) des Vereinigten Königreiches (Clarke und Tate 1999) könnte die Datenlage durch die Nutzung des Mikrozensus als Rotationspanel verbessert werden.

Der Mikrozensus ist sowohl eine Flächen- bzw. Klumpenstichprobe als auch eine rotierende Panelstichprobe, in der die Haushalte eines Auswahlbezirks vier Jahre lang befragt werden. Jedes Jahr wird ein Viertel der Auswahlbezirke (Klumpen) ausgetauscht. Aufgrund des thematisch breit gefächerten Fragenprogramms, des großen Stichprobenumfangs und der fast vollständigen Ausschöpfung bieten Mikrozensusdaten vielfältige Auswertungsmöglichkeiten. Seit dem Mikrozensusgesetz 1996 ist die rechtliche Voraussetzung gegeben, die Querschnittserhebungen des Mikrozensus zu einem Paneldatensatz zusammenzuführen. Auf Basis der Paneldaten ab 1996 ist es erstmals möglich, mit dem Mikrozensus Verläufe zu untersuchen. Damit eröffnet sich eine Reihe neuer Analysemöglichkeiten, beispielsweise von Analysen des familialen Wandels und von Berufs- und Bildungsverläufen. Da jedoch aufgrund des Prinzips der Flächenstichprobe wegziehende Personen im Mikrozensus nicht weiter befragt werden, können Verläufe nur für Personen ermittelt werden, die den Haushalt des Auswahlbezirks nicht verlassen. Somit hängt die Validität der Ergebnisse entscheidend davon ab, in welcher Weise die Panelausfälle mit den interessierenden Übergängen zusammenhängen. Diese Arbeit beschäftigt sich mit der Stichprobenselektivität bei der Analyse von Bildungsverläufen vom Besuch der gymnasialen Oberstufe bis zum Abitur

sowie von beruflichen Ausbildungen im dualen System bis zum Ausbildungsabschluss und dem Übergang in das Beschäftigungssystem.

Die Fragen potenzieller Verzerrungen infolge nicht erfasster räumlicher Mobilität sind im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) geförderten Verbundprojekts „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“ untersucht worden. Weitere Hauptaufgaben des gemeinsam von statistischen Ämtern und der Wissenschaft durchgeführten Projektes lagen in der Konstruktion und Evaluation von Gewichtungsvariablen zur Korrektur der Stichprobenselektivität. Um die Voraussetzungen für die Weitergabe der Paneldaten als Scientific Use File zu schaffen, waren außerdem ein Konzept zur faktischen Anonymisierung und ein Filekonzept zu entwickeln.<sup>1</sup>

Das Risiko selektiver Ausfälle hängt davon ab, wie stark die inhaltlich interessierenden Variablen mit der räumlichen Mobilität korreliert sind. Das Selektivitätsrisiko kann also je nach Fragestellung unterschiedlich hoch sein, sodass zur Einschätzung der Aussagekraft von Panelanalysen des Mikrozensus verschiedene Themenbereiche zu berücksichtigen sind. Zugleich sind Subpopulationen mit unterschiedlicher räumlicher Mobilitätsneigung zu untersuchen, um für Analysen praktische Empfehlungen geben zu können. Diese strategischen Gesichtspunkte sind im Verbundprojekt arbeitsteilig organisiert worden. Aus inhaltlichen und methodischen Gründen konzentriert sich diese Arbeit auf bildungsstatistische Zusammenhänge für die unter 30-jährige Bevölkerungsgruppe.

In inhaltlicher Hinsicht ist festzustellen, dass bisherige Auswertungen der Querschnittsdaten des Mikrozensus entweder nur zur Bildungsbeteiligung oder zum Bildungsabschluss vorgenommen werden konnten. Auf der Basis der Paneldaten ist es erstmals möglich, Ausbildungsverläufe im allgemein bildenden und beruflichen Schulwesen innerhalb des durch das Erhebungsdesign vorgegebenen Beobachtungsfensters von vier Zeitpunkten bis zum Schulabschluss zu betrachten sowie die Chancen einer Erwerbstätigkeit im erlernten Beruf zu analysieren. Falls die Panelausfälle bei diesen Analysen ignorierbar oder korrigierbar sind, kann das Mikrozensuspanel dazu beitragen, das für Deutschland festgestellte Defizit an Längsschnittdaten zu verringern.

Unter methodischen Gesichtspunkten zeigten erste Auswertungen des Mikrozensuspanel 1996-1999 zur Selektivitätsproblematik (Heidenreich 2002), dass neben jungen Familien mit Kindern bis zu zehn Jahren und älteren Personen im Alter ab 70 Jahren die Altersgruppe der 18- bis 30-Jährigen überdurchschnittlich räumlich mobil ist. Insbesondere im Alter von etwa 20 bis 25 Jahren verlassen viele Jugendliche den weiter befragten Haushalt ihrer Eltern, zum Beispiel aus ausbildungs- oder berufsbedingten Gründen. Zur Frage der Zuverlässigkeit der Ergebnisse des Mikrozensuspanels stellen deshalb Analysen dieser Teilpopulation besonders hohe Anforderungen.

## 1.1 Aufgabenstellung der Arbeit

Einerseits sind bei Jugendlichen und jungen Erwachsenen als eine räumlich hoch mobile Teilpopulation bei Analysen allgemeiner und beruflicher Ausbildungsetappen und des Übergangs vom Ausbildungs- in den Erwerbsstatus selektive Ausfälle zu erwarten. Andererseits ziehen beispielsweise nicht alle Abiturienten bei Beginn eines Studiums von zu Hause aus und auch Auszubildende verlassen oft nicht sofort nach Abschluss der Ausbildung die bisherige Wohnung. Im Unterschied zu Ausfallanalysen bei Querschnittsdaten liegen des Weiteren bei Paneldaten günstigere

<sup>1</sup> Für weitere Informationen siehe die WWW-Seiten des Verbundprojekts unter der Adresse <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Einfuehrung,?templateId=renderPrint.psml>.

Ausgangsbedingungen vor, weil Angaben zum Zeitpunkt vor dem Ausfall bekannt sind und gegebenenfalls für die Schätzung von Ausfallkorrekturen verwendet werden können. Die vermuteten Selektivitätsprobleme sind deshalb empirisch zu untersuchen. Diese methodischen Fragen der Verwendbarkeit des Mikrozensuspanels für Längsschnittanalysen stellen sich insbesondere vor dem Hintergrund gravierender Lücken längsschnittlicher Informationen in der empirischen Bildungsforschung (Weishaupt und Fickermann 2001; Weißhuhn 2001).

In Bezug auf Bildungsverläufe liegen zu den Übergängen der Schüler zwischen den einzelnen Schulformen des allgemein bildenden und beruflichen Schulwesens sowie zu Übergängen nach Abschluss der Erst- oder Zweitausbildung im allgemein bildenden und beruflichen Schulsystem praktisch keine Angaben vor. Die wenigen Studien der empirischen Bildungsforschung beziehen sich meist nur auf eine regional begrenzte Auswahl bestimmter Schulformen und sind teilweise veraltet (Bellenberg 1999). Die Daten der amtlichen Bildungsstatistik sind als Querschnitts- und Anstaltsbefragungen nicht in der Lage über Schulkarrieren Aussagen zu machen. Das heißt, Kontinuitäten und Brüche, Schulwechsel und Wiederholungen oder Chancen eines erfolgreichen Schulbesuchs sind nicht ausreichend darstellbar. Für die Arbeitsmarkt- und Bildungsforschung sowie die Sozialstrukturanalyse ist das Mikrozensuspanel vor allem interessant, weil nun damit Bildungs- und Erwerbsverläufe betrachtet werden können. Ob und mit welcher Güte das Mikrozensuspanel „dynamische Kennziffern“ (Weißhuhn 2001: 65) zur Verwertbarkeit von Bildung auf dem Arbeitsmarkt liefern kann, die unter anderem für die Analyse des Wandels der Berufs- und Wirtschaftsstruktur benötigt werden, hängt ganz entscheidend von den selektiven Ausfällen ab. Es ist deshalb zu klären, in welchem Maße durch räumliche Mobilität bedingte Verzerrungen vorliegen und ob diese gegebenenfalls korrigiert werden können. Für die Evaluation des Analysepotenzials können weitere Panelinformationen genutzt werden. Insofern nämlich auf die Individualangaben früherer Erhebungen zurückgegriffen werden kann, gestattet die Längsschnittverknüpfung neue Analysen zur zeitlichen Konsistenz der Antworten und damit zur Datenqualität.

## 1.2 Konzeption und Aufbau

Die vorliegende Arbeit behandelt die Analysemöglichkeiten von Bildungsverläufen am Beispiel des Besuchs der gymnasialen Oberstufe bis zum Abitur und der Auszubildenden des dualen Systems und ihren Übergängen ins Erwerbssystem nach dem Abschluss der Ausbildung. Für die Bearbeitung dieser Fragestellungen werden substanzwissenschaftliche Theorien benötigt. Diese Ansätze orientieren sich an Hypothesen der soziologischen Ungleichheits- und Mobilitätsforschung sowie der Humankapitaltheorie und anderen Arbeitsmarkttheorien. Die Theorien beziehen sich auf Merkmalszusammenhänge für die Gesamtdaten ohne selektive Ausfälle aufgrund unbeobachteter räumlicher Mobilität zu berücksichtigen.

In der interessierenden Altersgruppe ist die räumliche Mobilität eng mit dem Auszug aus dem Elternhaus verbunden. Für die Beantwortung des methodischen Kernproblems, ob Analysen des Mikrozensuspanels infolge des Ausfalls räumlich mobiler Haushalte und Personen zu verzerrten Ergebnissen führen (Ausgangshypothese), sind daher aus einer inhaltlichen Perspektive allgemein Migrationstheorien von Bedeutung. Das hierunter fallende Thema „Auszug aus dem Elternhaus“ liefert die für ein Mobilitätsmodell benötigten Informationen und ist deshalb den bildungsstatistischen Analysen konzeptionell vorgelagert.

Methodisch lässt sich die notwendige Verbindung zwischen dem Modell für die Gesamtdaten und dem Ausfallmodell durch ein statistisches Verfahren herstellen, in dem die gemeinsame Wahrscheinlichkeit der substanzwissenschaftlich interessierenden Variablen sowie den zur Er-



klärung von Ausfällen herangezogenen Variablen betrachtet wird. Mit einem solchen Modell können nicht nur Hypothesen zur Beziehungsstruktur zwischen den Analysevariablen, sondern auch Annahmen über Zusammenhänge zwischen dem Ausfall und den Analysevariablen untersucht werden.

Mit Ausnahme vollständig zufälliger Ausfälle, d. h. wenn die räumliche Mobilität weder mit der abhängigen noch mit den erklärenden Variablen zusammenhängt, lassen sich lediglich auf Basis der beobachteten Daten weitere Ausfallannahmen nicht prüfen. Für Überlegungen über die Plausibilität verschiedener Ausfallannahmen werden deshalb oft Sensitivitätsanalysen vorgenommen, mit denen die Variabilität oder Unsicherheit von Ausfallannahmen zum Ausdruck gebracht wird. Letztlich sind aber zur Validierung der Ausfallmodelle fundierte a priori Kenntnisse über die Ausfälle bzw. Zusammenhänge in der Grundgesamtheit notwendig. Für diesen Zweck werden deshalb externe Daten benötigt.

Im folgenden Kapitel werden die Daten des Mikrozensuspanel 1996-1999 und der Referenzdaten vorgestellt. Die Beschreibung des Mikrozensuspanel bezieht sich auf den für diesen Zeitraum gültigen Stichprobenplan und das Frageprogramm des Mikrozensus ab 1996 sowie die Zusammenführung der Querschnittserhebungen zum Rotationspanel und das verwendete Arbeitsfile des Verbundprojektes. Im Anschluss daran werden die als Referenzstatistiken verwendeten Daten der Schulstatistik und der Beschäftigtenstichprobe skizziert.

Das dritte Kapitel enthält die methodischen Grundlagen für die Aufdeckung und Behandlung selektiver Ausfälle. Im ersten Abschnitt wird zunächst die Klassifikation verschiedener Ausfalltypen vorgestellt. Nach einem Überblick zum möglichen Umgang mit Ausfällen bei kategorialen Daten im Abschnitt 3.2 konzentrieren sich die Abschnitte 3.3 und 3.4 auf statistische Modelle für diesen Datentyp der im Mikrozensus überwiegend vorliegenden Variablen und erläutern diese Modelle anhand einfacher Beispiele.

Das vierte Kapitel befasst sich mit Analysen zum Auszug aus dem Elternhaus. Zunächst werden theoretische Ansätze zum Auszug aus dem Elternhaus als Spezialfall von Migrationstheorien vorgestellt (4.1). Darauf aufbauend werden theoriegeleitet soziodemografische Angaben der Jugendlichen und ihres Elternhauses ausgewählt, die den Auszug beschleunigen oder verzögern. Im zweiten Abschnitt werden erste deskriptive Analysen berichtet. Der dritte Abschnitt enthält eine Beschreibung des für die Verlaufsdatenanalysen gewählten Modells sowie damit erzielte Ergebnisse. Abschließend werden die Ergebnisse zusammengefasst und vorläufige Schlussfolgerungen zur Stichprobenselektivität gezogen, die für die nachfolgenden bildungsstatistischen Analysen wichtige Vorinformationen über potenzielle Verzerrungen liefern.

Das fünfte Kapitel behandelt Bildungsverläufe vom Besuch der gymnasialen Oberstufe bis zum Abitur. Aufbauend auf den Ergebnissen zum Auszug aus dem Elternhaus werden für diese Subpopulation spezifische Ausfallhypothesen, theoretische Annahmen zum Zusammenhang zwischen Bildungsbeteiligung bzw. Bildungserfolg und soziodemografischen Variablen sowie sozialer Herkunft formuliert (5.1). Nach der Beschreibung der Abgrenzung der Analysestichprobe und Operationalisierung der Analysevariablen werden erste Ergebnisse des mit dem Mikrozensuspanel ermittelten Abiturabschlusses vorgestellt. Der dritte Abschnitt enthält deskriptive Vergleiche der Schüler- und Abiturientenzahlen des Mikrozensuspanels mit den Referenzdaten der Bildungsstatistik. In diesem Abschnitt werden insbesondere auch Fragen zur Datenqualität bzw. Konsistenz der Bildungsangaben aufgegriffen, die mit den Paneldaten erstmals untersuchbar sind. Anschließend folgen Ergebnisse multivariater Analysen zum Ausfall bzw. Wegzug im Mikrozensuspanel. Das Kapitel schließt mit einer Zusammenfassung der Befunde.

Im sechsten Kapitel werden berufliche Ausbildungsverläufe bis zum Übergang ins Erwerbsleben untersucht. Als Referenzdaten dienen hier die Daten der Beschäftigtenstichprobe. Der erste

Abschnitt befasst sich mit der Konstruktion vergleichbarer Abgrenzungen der Auszubildenden und Ausbildungsabschlüsse. Im zweiten Abschnitt folgen zunächst weitere deskriptive Vergleiche der Verteilungen der Auszubildenden, der Ausbildungsabschlüsse und -abbrüche im Mikrozensuspanel und in der Beschäftigtenstichprobe. Mit Bezug auf Ergebnisse zum Auszug aus dem Elternhaus und spezifischen, für diese Gruppe der Auszubildenden formulierten Hypothesen zum Zusammenhang zwischen Ausbildungserfolg und Ausfall werden anschließend statistische Analysen zum Ausfall vorgestellt. Der dritte Abschnitt enthält Tests zur Einsatzmöglichkeit von Gewichtungsfaktoren für die Korrektur von Verzerrungen. Deskriptive Analysen zu den Chancen der Beschäftigung im erlernten Beruf finden sich im vierten Abschnitt. Abschließend werden die Befunde zusammengefasst.

Im siebten Kapitel werden die einzelnen Ergebnisse zur Stichprobenselektivität und die Möglichkeiten und Grenzen der Analyse von Bildungsverläufen mit dem Mikrozensuspanel zusammenfassend diskutiert. Über das gegenwärtige Mikrozensuspanel 1996-1999 hinausgehend werden abschließend weitere Fragen thematisiert, mit welchen Mitteln das Potenzial des Mikrozensus für die Längsschnittanalyse vergrößert werden könnte.



## 2 Datenbeschreibung

In diesem Kapitel werden die Datengrundlagen des Mikrozensuspanels 1996-1999, das aus der Zusammenführung der Querschnittserhebungen entstanden ist, sowie die Referenzstatistiken beschrieben, die für die Überprüfung der Selektivität im Mikrozensuspanel verwendet werden. Der erste Abschnitt behandelt zunächst die zentralen Stichprobeneigenschaften des Mikrozensus und des Frageprogramms für den Zeitraum des Mikrozensuspanels (2.1.1). Anschließend werden die Konstruktion und der Aufbau der Paneldaten dargestellt (2.1.2). In den weiteren Abschnitten werden kurz die amtliche Bildungsstatistik (Abschnitt 2) und die IAB-Regionalstichprobe 1975-2001 (IABS-R01; Abschnitt 3) beschrieben. Die Fragen vergleichbarer Abgrenzungen zwischen Mikrozensuspanel und den Referenzstatistiken werden in den nachfolgenden Kapiteln ausführlich diskutiert.

### 2.1 Mikrozensuspanel 1996-1999

Der Mikrozensus wird seit 1957 in Deutschland als amtliche Haushaltsstichprobe über die Bevölkerung und den Arbeitsmarkt durchgeführt. Das durch entsprechende Gesetze vorgegebene Erhebungsprogramm enthält eine Vielzahl von demografischen, sozialen und wirtschaftlichen Angaben der Personen und Haushalte (vgl. Emmerling und Riede 1997; Lüttinger und Riede 1997). Für die ausgewählten zu befragenden Personen und Haushalte besteht eine gesetzliche Auskunftspflicht. Bis einschließlich 2004 fand die Befragung einmal im Jahr statt. Die Berichtswche lag i. d. R. Ende April eines Jahres. Seit 2005 wird der Mikrozensus kontinuierlich mit einer gleichmäßigen Verteilung der Interviews über das ganze Jahr erhoben.<sup>2</sup>

#### 2.1.1 Stichprobenplan und Erhebungsprogramm

Der Mikrozensus ist sowohl eine einstufig geschichtete Klumpen- bzw. Flächenstichprobe mit einem Auswahlatz von einem Prozent der Haushalte und Personen als auch ein Kurzzeitpanel mit partieller Rotation. Die Grundausswahl der gegenwärtig genutzten Stichproben erfolgte 1990 nach einem neuen Ziehungsverfahren. Die Primäreinheiten sind Auswahlbezirke (Klumpen), die i. d. R. aus benachbarten Wohnungen bestehen, welche auf Basis von Ergebnissen der Volkszählung von 1987 bzw. des Zentralen Einwohnerregisters von 1990 in den neuen Bundesländern gebildet wurden. Die durchschnittliche Klumpengröße liegt bei etwa neun Wohnungen. Alle Haushalte und Personen, die in den ausgewählten Klumpen wohnen, werden als (sekundäre) Erhebungseinheiten befragt. Zur Aktualisierung der Stichprobe um die seit der Grundausswahl neu entstandenen Wohnungen werden diese auf Basis der Bautätigkeitsstatistik in einer sogenannten Neubauschicht zu separaten Auswahlbezirken zusammengefasst. Die Primäreinheiten sind regional nach dem Bundesland und über 200 Regionalschichten sowie fachlich zur Berücksichtigung unterschiedlicher Siedlungsstrukturen und Wohnformen nach der Gebäudegrößenklasse geschichtet (Heidenreich 2002; Meyer 1994; Rendtel und Schimpl-Neimanns 2001). Aufgrund der Auskunftspflicht liegt die Teilnahmequote der Haushalte bei rund 97 Prozent.

---

2 Mit dem Mikrozensusgesetz 2005 ergaben sich weitere Modifikationen des Erhebungsprogramms. Auf diese Neuerungen wird im Folgenden nicht eingegangen (siehe dazu Afentakis und Bihler 2005; Lotze und Breiholz 2002a, b).

Die Auswahlbezirke und die darin wohnenden Haushalte und Personen verbleiben vier Jahre lang in der Befragung und bilden ein sogenanntes Rotationsviertel. Jährlich wird ein Viertel der Auswahlbezirke ausgetauscht, d. h. es liegen für maximal vier Jahre Verlaufsangaben vor. Der Mikrozensus ist somit eine Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten (partielle Rotation).

Der hauptsächliche Grund für die Wahl dieses Designs der partiellen Rotation besteht darin, die hohen Kosten einer jährlich neuen Stichprobenziehung zu reduzieren und den Aufwand der Feldarbeit zu verringern. Ein weiterer methodischer Vorteil ergibt sich daraus, dass durch die Überlappung von Erhebungseinheiten der Stichprobenfehler der Differenz von Populationsschätzungen verringert wird; z. B. beim Vergleich von Gesamtwerten zweier Erhebungsjahre. Aber schon bei der Planung zur Einführung des Mikrozensus in Deutschland waren Wiederholungsbefragungen auch zur Erfassung von Veränderungen erwogen worden. (Koller und Herberger 1960).<sup>3</sup> Durch die Einführung der partiellen Rotation ab Oktober 1962 wurden diese Auswertungen prinzipiell möglich, jedoch in den Folgejahren insgesamt wenig genutzt. Sie waren schließlich ab 1982 infolge von Löschungsvorschriften zur Längsschnittidentifikation der Stichprobeneinheiten nicht mehr realisierbar. Dies ist erst wieder mit dem Mikrozensusgesetz 1996 rechtlich möglich geworden.

Die in Rotationsviertel aufgeteilte einprozentige Stichprobe wird weiter unterteilt, um Stichproben mit reduziertem Auswahlsatz durchführen zu können. Dies wird u. a. genutzt, um Anforderungen der in den Mikrozensus als Substichprobe integrierten Arbeitskräfteerhebung der EU (Labour Force Survey) zu erfüllen. Aufgrund von Vorgaben über die Größe des Stichprobenfehlers bestimmter Merkmale für Regierungsbezirke beträgt der Auswahlsatz dieser Unterstichproben in den Querschnittsdaten 0,4, 0,6, 0,8 oder ein Prozent in Abhängigkeit von der Regierungsbezirksgröße. Im Bundesdurchschnitt liegt der disproportionale Auswahlsatz bei rund 0,45 Prozent.

Die Fallzahlen des Mikrozensus werden in einem zweistufigen Verfahren hochgerechnet. Auf der ersten Stufe erfolgt eine Korrektur des Unit-Nonresponse von rund drei Prozent für die nicht erreichten Haushalte. Auf der zweiten Stufe werden die nach dieser Ausfallkorrektur gewichteten Mikrozensusergebnisse an bekannte demografische Randverteilungen angepasst. Diese Anpassung an Ergebnisse der laufenden Bevölkerungsfortschreibung erfolgt für die Merkmale Geschlecht und Staatsangehörigkeit (Deutsche vs. Ausländer) auf der regionalen Ebene von sogenannten Anpassungsschichten sowie für Soldaten und Wehrpflichtige an entsprechende Bestandsmeldungen auf Regierungsbezirksebene (siehe Heidenreich 1994). Die Anpassungsschichten umfassen durchschnittlich mindestens 500.000 Einwohner.

Der Schwerpunkt des Fragenprogramms liegt auf erwerbsstatistischen und soziodemografischen Angaben. Daneben werden auch Fragen zum Lebensunterhalt und dem Einkommen, zum allgemein bildenden und beruflichen Abschluss, zur Wohnsituation oder zur Gesundheit gestellt. Jedoch werden nicht alle diese Themen mit dem vollen Auswahlsatz von einem Prozent und nicht in jedem Jahr erhoben. Das Fragenprogramm unterscheidet sich nach einem Grundprogramm sowie einem Ergänzungs- und Zusatzprogramm. Die zum Grundprogramm gehörenden Fragen (z. B. Erwerbstätigkeit, Familien- und Haushaltszusammenhang und Staatsangehörigkeit) werden jährlich mit vollem Auswahlsatz gestellt. Das Ergänzungsprogramm wird ebenfalls jährlich, jedoch nur für eine Substichprobe von bis zu 0,5 Prozent durchgeführt; hierzu zählen insbesonde-

---

3 Auf die Weiterbefragung umgezogener Haushalte wurde jedoch aus organisatorischen Gründen verzichtet, da die neuen Adressen nur aufwendig und unvollständig zu ermitteln waren (Herberger 1973: 63).

re Fragen der EU-Arbeitskräftestichprobe (z. B. Aus- und Fortbildung, frühere Erwerbstätigkeit, Retrospektivfragen zur Erwerbstätigkeit und dem Wohnsitz ein Jahr zuvor). Dagegen werden die Zusatzprogramme nur im Abstand von vier Jahren durchgeführt. Sie können teilweise mit vollem Auswahlsatz oder nur für eine 0,5-Prozent-Substichprobe erhoben werden (z. B. Zusatzangaben für Ausländer, Gesundheit, berufliche Bildung und allgemeine Weiterbildung).

Die meisten Fragen sind auskunftspflichtig. Daneben gibt es weitere Fragen und Themenbereiche, die von der Auskunftspflicht ausgenommen sind. Hierzu zählen beispielsweise Fragen der EU-Arbeitskräftestichprobe, Fragen zur Gesundheit sowie (für den Zeitraum zwischen 1996 und 2004) die Angaben zum allgemeinen und beruflichen Abschluss von Personen über 50 Jahren.

Hinsichtlich der Befragungsmethode dominiert die persönliche Befragung. Rund 15 Prozent der Interviews basieren auf schriftlichen Auskünften der Befragten. Fremdauskünfte (Proxy-Interviews) sind möglich, z. B. wenn Eltern die Fragen zum Schulbesuch und Bildungsabschluss ihrer Kinder beantworten. Erst ab der Erhebung von 1999 wird im Rahmen der Unterstichprobe des Mikrozensus erfasst, ob es sich um Selbst- oder Fremdauskünfte handelt. Insgesamt beruhen rund 30 Prozent der Angaben auf Proxy-Interviews (Breiholz 2000). Bis zum Mikrozensus von 2004 bezieht sich der Berichtszeitraum i. d. R. auf die letzte feiertagsfreie Woche im April eines Jahres (Berichtswoche).

## 2.1.2 Zusammenführung der Querschnittsangaben

Für die Konstruktion des Mikrozensuspanels sind, wie oben bereits angesprochen, die partielle Rotation sowie das Prinzip der Flächenstichprobe von zentraler Bedeutung. Grundsätzlich lassen sich die Rotationsviertel der Mikrozensusstichprobe zu verschiedenen Panels verbinden. Abbildung 2.1 zeigt, wie sich die Mikrozensusquerschnitte aus den Rotationsgruppen der einzelnen Stichproben zusammensetzen. Beispielsweise wurden 1996 aus der dritten Stichprobe die Rotationsgruppen drei und vier und aus der vierten Stichprobe die Rotationsgruppen eins und zwei befragt. Durch Kombination verschiedener Stichproben und Rotationsgruppen können also Längsschnittdaten konstruiert werden, die Zeiträume von zwei, drei oder vier Jahren abdecken (siehe Abb. 2.2).

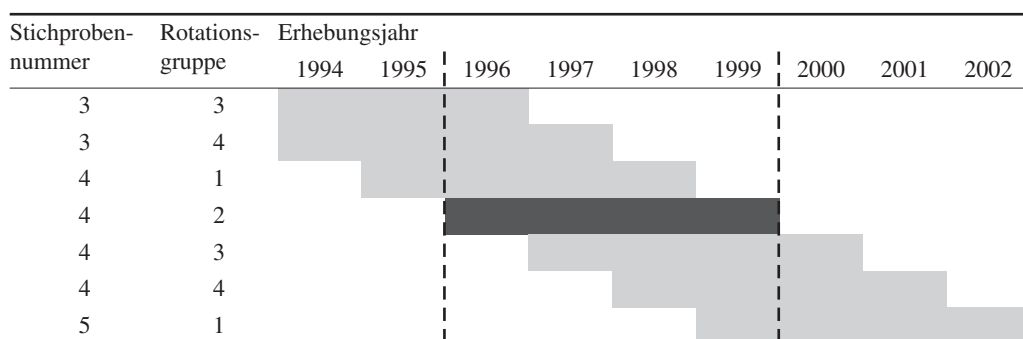


Abbildung 2.1: Rotation im Mikrozensus

Weil die aus dem Auswahlbezirk wegziehenden Haushalte und Personen nicht weiterbefragt werden, reduziert sich das Zusammenführungspotenzial um die Fort- bzw. Zuzüge sowie Sterbefälle und Geburten (siehe Abb. 2.3).

Beobachtungsdauer	Periode	Stichprobe / Rotationsgruppe
4 Jahre	1996 - 1999	4/2
3 Jahre	1996 - 1998	4/2, 4/1
	1997 - 1999	4/2, 4/3
2 Jahre	1996 - 1997	4/2, 3/4, 4/1
	1997 - 1998	4/2, 4/3, 4/1
	1998 - 1999	4/2, 4/3, 4/4

Abbildung 2.2: Beobachtungsdauer, Zeitraum und Zusammensetzung der Stichprobe

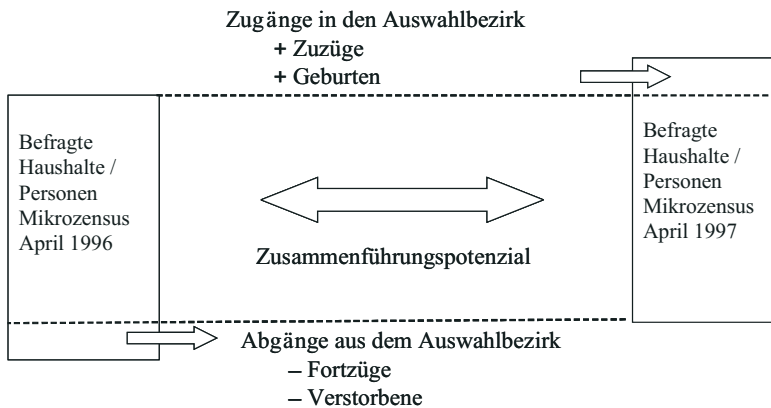


Abbildung 2.3: Systematik der Zusammenführung zweier Querschnitte

Für die Zusammenführung kann das hierarchisch aufgebaute System von Ordnungsnummern genutzt werden, mit denen die Zugehörigkeit von Personen, Haushalten und Wohnungen in einem Auswahlbezirk festgelegt ist. Zwar wurde es mit dem Mikrozensusgesetz 1996 wieder möglich, über die vier Erhebungszeitpunkte identische Ordnungsnummern zu bilden und zu speichern. Bei den ersten Versuchen der Zusammenführung zeigten sich jedoch Probleme aufgrund von im Zeitverlauf inkonsistenten Ordnungsnummern. Nur mithilfe der Merkmale Geburtsjahr und Geschlecht, die zusätzlich zu den für die Zusammenführung benutzten Identifikatoren (Bundesland, Regierungsbezirk, Auswahlbezirk und Haushalt) verwendet wurden, konnten auf Personenebene – zum Teil aber erst nach umfangreichen manuellen Korrekturen – circa 97 Prozent der Sätze verknüpft werden. Auf Haushaltsebene konnten rund 90 Prozent der Haushalte zusammengeführt werden (Heidenreich 2002; Herter-Eschweiler 2003: 50-53; Statistisches Bundesamt 2006a). Nicht zusammengeführt werden konnten Personen in Gemeinschaftsunterkünften (u. a. Alten- oder Studentenwohnheime), da ohne eine längsschnittliche Personennummer das Geburtsjahr und das Geschlecht in diesen Einrichtungen aufgrund zu starker Klumpungen nicht zur Identifikation der zu einer Person gehörenden Querschnittsangaben geeignet sind.

Das hier verwendete Panelfile basiert auf den vom Statistischen Bundesamt zusammengeführten Querschnittserhebungen des von 1996 bis 1999 befragten Rotationsviertels. Das Arbeitsfile des Methodenverbundprojekts, mit dem die Fragen der Stichprobenselektivität untersucht werden, ist eine faktisch anonymisierte 60-Prozent-Substichprobe von Auswahlbezirken, wobei alle

Haushalte und Personen eines ausgewählten Auswahlbezirks unabhängig vom Zusammenführungserfolg enthalten sind. Bezogen auf Privathaushalte entspricht das Arbeitsfile circa einer 65-Prozent-Haushaltssubstichprobe. Der Stichprobenumfang beträgt rund 54.500 Haushalte und 120.000 Personen pro Erhebungszeitpunkt.<sup>4</sup>

Die Übernahme aller Einheiten eines für die Substichprobe ausgewählten Auswahlbezirks bietet wichtige Vorteile: Aufgrund des Prinzips der Flächenstichprobe werden in den Auswahlbezirken „automatisch“ alle Populationsveränderungen erfasst, sodass auch Querschnittsanalysen mit dieser Substichprobe realisierbar sind. Insbesondere geben Vergleiche räumlich mobiler und räumlich immobiler Gruppen wichtige Anhaltspunkte für Selektivitätsanalysen. Infolge von Ausfällen bzw. Weg- und Zuzügen sowie bei nicht zusammenführbaren Haushalten oder Personen liegen nicht für alle Einheiten Angaben für die Erhebungszeitpunkte von 1996 bis 1999 vor. Des Weiteren können in geringem Umfang Angaben für Erhebungszeitpunkte fehlen, falls die Haushalte für die Befragung nicht erreichbar sind (temporäre Ausfälle). Das Mikrozensuspanel ist somit ein unbalanciertes Panel.

Bis auf wenige Ausnahmen enthält das Arbeitsfile des Mikrozensuspanels alle erfragten Merkmale des Mikrozensus, einschließlich der Merkmale aus den Zusatz- und Ergänzungsprogrammen. Des Weiteren sind Identifikatoren bzw. Ordnungsnummern der Erhebungseinheiten sowie abgeleitete Variablen (z. B. zum Haushalts- und Familienkontext sowie Typisierungen) und methodische Variablen (Filter, Hochrechnungsfaktoren etc.) der Querschnittsdaten nutzbar.

Das Panelfile enthält allerdings aus Datenschutzgründen keine Unterstichprobenmerkmale für den Zeitpunkt 1999. Z. B. fehlt die 1999 erstmals erfasste Information, ob die Angaben auf Selbst- oder Fremdauskünften (Proxy-Angaben) beruhen. In diesem Zusammenhang ist auch anzumerken, dass die statistischen Ämter die Erhebung der Unterstichprobe so organisiert haben, dass einerseits eine möglichst gleichmäßige Belastung der betreffenden Haushalte über die vier Erhebungszeitpunkte und andererseits eine Entlastung bei der ersten und vierten Befragung erreicht wird. So werden die Fragen der EU-Arbeitskräfteerhebung bis auf wenige Ausnahmen in der vierten und letzten Erhebung nicht mehr gestellt. Längsschnittanalysen für Unterstichprobenmerkmale sind daher im Mikrozensuspanel nur eingeschränkt möglich.

Das Arbeitsfile enthält verschiedene GewichtungsvARIABLEN, die für die Hochrechnung und zur Korrektur von Verzerrungen, die infolge räumlicher Mobilität auftreten, verwendet werden können (Statistisches Bundesamt 2006a: 91ff.). Zum einen enthält das File die regulären Hochrechnungsfaktoren, die bei der sogenannten gebundenen Hochrechnung verwendet werden. Diese Gewichte gewährleisten bei den Querschnittsdaten in Kombination mit der Designgewichtung mit dem Kehrwert der Inklusionswahrscheinlichkeit eine Anpassung ausgewählter Merkmale der Stichprobe an entsprechende Verteilungen der Grundgesamtheit bzw. der laufenden Bevölkerungsfortschreibung. Ergänzend wurden Gewichte gebildet, welche die Ziehungswahrscheinlichkeiten des Mikrozensuspanels und die Quoten erfolgreich zusammengeführter Privathaushalte berücksichtigen (Herter-Eschweiler 2003). Zum anderen enthält das Arbeitsfile die im Teilprojekt „Methoden“ entwickelten Hochrechnungsgewichte für Längsschnittdaten. Hierbei können drei Variablengruppen unterschieden werden: (1) Längsschnittgewichte resultierend aus einer simultanen Anpassung (Kalibrierung) der Stichprobendaten sowohl an Daten aus der laufenden Bevölkerungsfortschreibung als auch an demografische Veränderungsdaten zur Bevölkerung, wie Geburten, Eheschließungen und Ehescheidungen, Zuzüge aus dem Ausland, Sterbefälle und

4 Das nach Abschluss des Projektes als Scientific Use File verfügbare Mikrozensuspanel unterscheidet sich von dem Arbeitsfile unter anderem durch einen höheren Auswahlsatz von 70 Prozent (siehe Schimpl-Neimanns 2006a; Statistisches Bundesamt 2006a; Wirth 2006).



Fortzüge ins Ausland (Rohloff 2005). (2) Diese Längsschnittgewichte wurden ergänzt um Kehrwerte der Wahrscheinlichkeit der räumlichen Immobilität, die durch ein Logit-Modell geschätzt wurden (Basic et al. 2005). (3) Schließlich enthält das File die Mobilitätsgewichte des Logit-Modells. Die ersten beiden Gewichte (1+2) stehen nur für Längsschnitte auf Basis des Jahres 1996 zur Verfügung, d. h. für die Zeiträume 1996-1997, 1996-1998 und 1996-1999.

In den durch die Kombination regionaler und demografischer Merkmale gebildeten Anpassungsgruppen gilt die Annahme gleicher Antwort- bzw. Ausfallwahrscheinlichkeiten („response homogeneity groups“; siehe Särndal et al. 1997: 577ff.). Um mit diesen Gewichten die durch unbeobachtete räumliche Mobilität bedingte Verzerrung korrigieren zu können, muss in den Anpassungsgruppen die Annahme bedingt zufälliger Ausfälle (siehe Abschnitt 3.3.2) zutreffen. D. h. es sollte kein Zusammenhang zwischen dem interessierenden Merkmal und dem Ausfall bzw. der räumlichen Mobilität bestehen.

## 2.2 Amtliche Bildungsstatistik

Für Vergleiche der Besucher und Absolventen der gymnasialen Oberstufe des Mikrozensuspanels werden Referenzdaten der Bildungsstatistik verwendet. Die amtliche Schulstatistik wird als Anstaltserhebung durchgeführt und liefert Populationsdaten zum Besuch und Abschluss allgemein bildender Schulen. Sie basiert auf einer Vollerhebung und ist als Teil der amtlichen Bildungsstatistik in den einzelnen Bundesländern unterschiedlich organisiert. Durch die Orientierung an einem Minimalprogramm können aber Bundesergebnisse zusammengestellt werden (Statistisches Bundesamt 1996; Weishaupt und Fickermann 2001). Die Statistik beruht auf Meldungen der Schulleiter (Anstaltsbefragung). Räumlichen Darstellungen, wie beispielsweise dem Schulbesuch nach Bundesland, liegt der Schulort zugrunde. Die summarischen Angaben umfassen unter anderem die Zahl der Schüler nach Klassenstufe, Alter, Geschlecht und Staatsangehörigkeit. Dass Klassenstufe und Alter nicht in Kombination ausgewiesen werden, schränkt Vergleiche mit dem Mikrozensus ein. Zu beachten ist auch, dass zur Verringerung der Belastung der Auskunft gebenden Schulen der Schulbesuch nicht in jedem Jahr und nicht in jedem Bundesland differenziert erhoben wird. Beispielsweise werden die nach Geburtsjahr gegliederten Schülerzahlen teilweise nur im Abstand mehrerer Jahre ermittelt. Die Altersverteilungen der dazwischen liegenden Zeitpunkte werden auf Basis der ausführlichen Strukturserhebungen geschätzt.

Bei der Schulstatistik handelt es sich um Stichtagserhebungen, deren Berichtszeitraum sich jeweils auf den Beginn eines Schuljahres (Herbst) bezieht. Die Angaben über Schulentlassene nach Abschlussart beziehen sich auf das vergangene Schuljahr bzw. das Schuljahresende.

## 2.3 IAB-Regionalstichprobe 1975-2001

Für die Validierung des Mikrozensuspanels in Bezug auf die Fragestellung der Übergänge von der beruflichen Ausbildung bis zur Ausübung des erlernten Berufs nach dem Ende der Ausbildung kommt insbesondere die Beschäftigtenstatistik in Frage. Die Beschäftigtenstatistik bzw. die sogenannte Historikdatei des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB) enthält Angaben der Längsschnittpopulation sozialversicherungspflichtig beschäftigter Personen; dies sind u. a. alle Auszubildenden des dualen Systems. Da Erwerbsverläufe von Auszubildenden auf Basis der Beschäftigtenstatistik nicht mit dem Problem der Panelmortalität behaftet sind, kann der Vergleich des Mikrozensuspanels mit der Beschäftigtenstatistik zur Aufdeckung des Selektionsbias im Mikrozensuspanel beitragen.

Die hier verwendete IAB-Regionalstichprobe (IABS-R01; im Folgenden kurz als Beschäftigtenstichprobe bezeichnet) enthält für den Zeitraum von 1975 bis 2001 vollständige Verlaufsangaben zur Erwerbstätigkeit und zu Lohnersatzleistungen für eine Stichprobe von zwei Prozent (Hammann et al. 2004).<sup>5</sup>

Ausgangsbasis der Stichprobenziehung war die sogenannte Beschäftigten- und Leistungsempfängerhistorik des IAB. In dieser Datei sind Meldungen zu sozialversicherungspflichtigen Beschäftigten in der Beschäftigtenstatistik sowie Meldungen über den Bezug von Lohnersatzleistungen der Bundesagentur für Arbeit (Arbeitslosengeld und Arbeitslosenhilfe sowie Unterhaltsgeld) zusammengefasst.

Grundlage der Beschäftigtenstatistik ist das Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, das von den Arbeitgebern für alle sozialversicherungspflichtig Beschäftigten Meldungen an die Sozialversicherungsträger verlangt. Ausgenommen davon sind u. a. Beamte und Selbstständige. Geringfügig Beschäftigte und mithelfende Familienangehörige sind erst seit April 1999 eingeschlossen. Die Beschäftigtenhistorik des IAB umfasst gegenüber den Stichtags- und Jahreszeitraumdaten der Beschäftigtenstatistik auch verspätet gemeldete Beschäftigungsverhältnisse; seit 1998 werden Daten nach 18 Monaten Wartezeit verwendet. Daneben sind in der Beschäftigtenhistorik Datenbereinigungen vorgenommen worden. Es handelt sich somit nicht um Umfragedaten, sondern um Angaben, die im Laufe des Verwaltungshandelns entstehen (sogenannte prozessproduzierte Daten). Die Daten enthalten im Vergleich zum Mikrozensus und zu anderen Umfragedaten nur wenige soziodemografische und sozioökonomische Variablen. Sofern es sich um meldepflichtige Merkmale wie Bruttoentgelt oder Beschäftigungszeiten handelt, ist von einer sehr hohen Datenqualität auszugehen. Bei anderen Merkmalen (z. B. Beruf oder Bildungsabschluss), die nicht versicherungsrechtlichen, sondern statistischen Zwecken dienen, ist wie bei Umfragedaten mit Codier- und Klassifikationsfehlern zu rechnen. Ein weiterer wichtiger Unterschied zum Mikrozensuspanel besteht darin, dass die Regionalangaben der Beschäftigtenhistorik sich nicht auf den Wohnort, sondern auf den Ort der Arbeitsstätte beziehen.

Die Stichprobe ist eine geschichtete einstufige Zufallsauswahl aller sozialversicherungspflichtig Beschäftigten, die von 1975 bis 2001 wenigstens eine Beschäftigungsmeldung hatten. Geschichtet wurde nach der Variablenkombination Staatsangehörigkeit (Deutsche / Nicht-Deutsche) und Bundesgebiet (Alte / Neue Bundesländer) mit vier Schichten (proportionale Aufteilung). Die Stichprobe enthält für rund 1.300.000 Personen über 21.000.000 Beschäftigungsmeldungen.

Aufgrund der in der Beschäftigtenstichprobe tagesgenau berichteten Beschäftigungszeiten können Statusangaben für die im Mikrozensus erfassten Berichtswochen ermittelt werden, sofern in diesem Zeitraum eine sozialversicherungspflichtige Beschäftigung oder ein Leistungsempfang vorliegt. In der Beschäftigtenstichprobe sind auch Mehrfachbeschäftigungen ausgewiesen. Beispielsweise kann neben einer Hauptbeschäftigung noch eine weitere Beschäftigung ausgeübt werden. Falls mehrere Meldungen vorliegen, wird im Folgenden aus Gründen der Vergleichbarkeit mit dem Mikrozensuspanel nur die jeweilige Hauptbeschäftigung ausgewählt.

---

5 Ausführliche Quellenangabe: Die Datengrundlage dieses Beitrages bildet die faktisch anonymisierte IAB-Regionalstichprobe 1975-2001 (IABS-R01). Dabei handelt es sich um eine Zwei-Prozent-Stichprobe aller sozialversicherungspflichtig Beschäftigten ergänzt um Zeiten des Leistungsbezugs. Grundlage der Stichprobenziehung ist die Beschäftigten-Leistungsempfänger-Historik (BLH) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit. Die IABS-R01 kann über das Zentralarchiv für Empirische Sozialforschung (Köln) bezogen werden. Für die Verwendung der Daten in diesem Beitrag trägt das IAB keine Verantwortung.



### 3 Methodische Grundlagen

Ziel dieses Kapitels ist es, die methodischen Grundlagen für den Umgang mit dem Kernproblem der durch die aus dem Haushalt des Auswahlbezirks wegziehenden Personen entstehenden Ausfälle im Mikrozensuspanel darzustellen. Das einfachste Vorgehen könnte darin bestehen, bei Längsschnittanalysen nur die Daten von räumlich Immobilen zu verwenden. Wegen des geringen Stichprobenumfangs führt dies bei Populationsschätzungen, wie z. B. der Zahl oder dem Anteil von Ausbildungsabsolventen, zu einem höheren Stichprobenfehler bzw. bei der Schätzung von Regressionsmodellen zu einem höheren Standardfehler der Koeffizienten. Gegebenenfalls unterscheidet sich aber die Umzugs- bzw. Ausfallwahrscheinlichkeit für bestimmte Teilgruppen. Wenn beispielsweise die Wahrscheinlichkeit eines Umzugs für Personen nach Erreichen des Ausbildungsabschlusses höher ist als für Personen, die sich noch in der Ausbildung befinden, oder wenn die Umzugswahrscheinlichkeit mit der zeitlichen Nähe zum Abschlusszeitpunkt korreliert, hängen die Ausfälle systematisch mit dem interessierenden Ereignis zusammen. Würde man in diesem Fall räumlich mobile Personen vernachlässigen, wäre die resultierende Stichprobe der räumlich Immobilen nicht mehr als einfache Zufallsauswahl der interessierenden Gesamtheit zu betrachten. Die daraus entstehenden Verzerrungen können gravierender sein als der eingangs genannte Präzisionsverlust, der durch die bloße Verringerung der Stichprobe entsteht.

Das Ausmaß der möglichen Verzerrung hängt allgemein vom Anteil der ausgefallenen Personen innerhalb der betrachteten Gruppe und der Kovariation der interessierenden Variablen mit dem Ausfall ab. Es ist deshalb vordringlich notwendig, diese Zusammenhänge zwischen dem Ausfall und den Analysevariablen genauer auf ihre Implikationen hinsichtlich möglicher Verzerrungen zu untersuchen. Dies geschieht im folgenden ersten Abschnitt. Die darauf aufbauende Darstellung im zweiten Abschnitt konzentriert sich auf die Schätzung statistischer Modelle. Anschließend wird anhand einer einfachen Beispieltabelle beschrieben, mit welchen Verfahren für die im Mikrozensus hauptsächlich vorliegenden qualitativen Variablen bzw. kategorialen Daten Ausfalluntersuchungen und modellbasierte Ausfallkorrekturen vorgenommen werden können. Abschließend wird gezeigt, wie mit diesen Modellen auch Referenzdaten zur Validierung genutzt werden können.

#### 3.1 Klassifikation der Ausfallprozesse

Die Klassifikation der Ausfallprozesse folgt der von Rubin (1976) entwickelten Systematik, mit der die Implikationen fehlender Werte bei inferenzstatistischen Schlüssen dargestellt wurden.

Bei fehlenden Werten in einer oder mehreren Variablen liegen Datenkonstellationen vor, die mittels Indikatorvariablen unterschieden werden können. Für Analyseeinheiten ohne fehlende Werte stehen vollständige Angaben zur interessierenden Variablen  $Y$  (z. B. Schulbesuch oder Ausbildungsabschluss) sowie zu einer oder mehreren Kovariaten  $X$  zur Verfügung. Die Indikatorvariable  $R$  (Response) zeigt für diese Gruppe den Wert Eins ( $R = 1$ ) an.<sup>6</sup> Für die Gruppe ausgefallener Personen mit fehlenden Angaben zu  $Y$  ( $R = 0$ ) liegt dagegen nur  $X$  vor.  $Y$  setzt sich somit aus beobachteten („observed“) und fehlenden („missing“) Daten zusammen:  $(Y_{obs}, Y_{mis})$ . Im Folgenden wird angenommen, dass  $X$  in jedem Fall beobachtet werden kann. Dies entspricht auch der Datenlage in den späteren Analysen (siehe Kapitel 5 und 6).

6 Der Ausfallindikator wird bei Little und Rubin (2002)  $M$  (missing) genannt. Die nachstehende Darstellung folgt auch Copeland (2004: 14f.).

Die gemeinsame Verteilung der Analysevariablen  $(X, Y)$  und des Ausfallindikators  $R$  kann für ein statistisches Modell, in dem  $\theta$  und  $\varphi$  zu schätzende Parameter des unter inhaltlichen Aspekten interessierenden Strukturmodells der Analysevariablen bzw. des Ausfallmodells bezeichnen, wie folgt faktorisiert werden:

$$f(X, Y, R | \theta, \varphi) = f(X, Y | \theta) f(R | X, Y, \varphi) \quad (3.1)$$

Die Ausfallfunktion auf der rechten Seite im zweiten Teil der Gleichung beschreibt die bedingte Wahrscheinlichkeit von Ausfällen, die abhängig von  $X, Y$  und  $\varphi$  sind.

Betrachtet man das Entstehen fehlender Daten als Zufallsexperiment und den Indikator  $R$  als Zufallsvariable, können hinsichtlich der Zusammenhänge zwischen dem Ausfall und anderen Variablen drei Situationen unterschieden werden (Little und Rubin 2002: 11f.): vollkommen zufällige, bedingt zufällige und nicht ignorierbare Ausfälle.

Falls die Wahrscheinlichkeit eines Ausfalls bzw. fehlenden  $Y$ -Wertes weder mit  $Y$  noch mit  $X$  zusammenhängt, gilt für alle  $Y, X$ - und  $\varphi$ -Werte:

$$f(R | X, Y, \varphi) = f(R | \varphi) \quad (3.2)$$

Diese Situation wird als vollkommen zufälliger Ausfall („missing completely at random“, MCAR) bezeichnet. Die ausgefallenen bzw. fortgezogenen Personen können dann als einfache Zufallsstichprobe aus der insgesamt zufällig ausgewählten Stichprobe angesehen werden. Im Fall von MCAR entsprechen die lediglich auf Basis der Daten ohne Ausfälle ermittelten Analysen zum Zusammenhang zwischen  $Y$  und  $X$  den Ergebnissen, die man bei Berücksichtigung aller Befragten erhalten würde.

Trifft die MCAR-Annahme (3.2) zu, vereinfacht sich die in Gleichung (3.1) formulierte Schätzung der interessierenden Zusammenhänge zwischen  $Y$  und  $X$  zu

$$f(Y, R | X, \theta, \varphi) = f(Y | X | \theta) f(R | \varphi).$$

Hängt der Ausfall von  $X$ , aber nicht von  $Y$  ab, gilt für die Gruppe mit fehlenden Werten in  $Y$  ( $Y_{mis}$ ):

$$f(R | X, Y, \varphi) = f(R | X, Y_{obs}, \varphi) \quad (3.3)$$

Dies entspricht dem Typ eines bedingt zufälligen Ausfalls („missing at random“, MAR). MAR trifft beispielsweise zu, wenn Wegzüge mit dem Alter korrelieren, aber in jeder Altersgruppe kein Zusammenhang zwischen Schulbesuch oder Ausbildungsabschluss und Ausfall besteht. In diesem Fall können die ausgefallenen Personen wie eine einfache Zufallsstichprobe aus der jeweiligen Unterstichprobe von Personen mit denselben  $X$ -Werten behandelt werden.

Unter dieser Bedingung (3.3) lässt sich die Schätzung der interessierenden Zusammenhänge in (3.1) als  $f(Y, R | X, \theta, \varphi) = f(Y | X, \theta) f(R | X, \varphi)$  darstellen. Wenn  $X$  beobachtet wurde, können bei vollständigem und bedingt zufälligem Ausfall die Antwortwahrscheinlichkeiten auf Grundlage der vorliegenden Daten leicht berechnet und für Ausfallkorrekturen, wie z. B. Gewichtung mit dem Kehrwert der geschätzten Antwortwahrscheinlichkeit (Horwitz-Thompson-Schätzfunktion), verwendet werden. In der Regel werden bei Paneldaten hierfür Angaben zum Zeitpunkt vor dem Ausfall herangezogen.

Treffen die MCAR- oder MAR-Annahmen zu, kann die gemeinsame Verteilung von  $Y, X$  und  $R$  in ein Strukturmodell und in ein Ausfallmodell faktorisiert werden. Sind die Parameter des Struktur- und Ausfallmodells  $(\theta, \varphi)$  unabhängig voneinander schätzbar (Separierbarkeit), können somit alle Informationen zum Zusammenhang zwischen  $Y$  und  $X$  in der Gesamtheit (inkl. Ausfäll-

len) aus den beobachteten Daten geschätzt werden. Diese beiden Ausfalltypen werden deshalb als ignorierbar bezeichnet.

Ist jedoch die Wahrscheinlichkeit eines Ausfalls mit der abhängigen Variablen verbunden, d. h. wenn

$$f(R | X, Y, \varphi) = f(R | X, Y_{obs}, Y_{mis}, \varphi) \text{ oder} \\ f(R | X, Y, \varphi) = f(R | X, Y_{mis}, \varphi) \quad (3.4)$$

zutrifft, spricht man von nicht ignorierbaren bzw. nicht zufälligen Ausfällen („non-ignorable nonresponse“ (NINR), bzw. „missing not at random“ (MNAR)). Der Zusammenhang zwischen dem Ausfall und der abhängigen Variablen muss dann wie im Unterabschnitt 3.3.3 beschrieben explizit modelliert werden. Bei Verlaufsanalysen bedeutet dies, dass der Beobachtungszeitraum und das Ereignis nicht unabhängig sind.

## 3.2 Selektionsmodelle für kategoriale Daten

Für die Modellierung der gemeinsamen Verteilung der Analysevariablen ( $X$ ,  $Y$ ) und des Ausfallindikators  $R$  werden hauptsächlich Selektionsmodelle und Pattern-Mixture Modelle herangezogen (Little und Rubin 2002).

Selektionsmodelle<sup>7</sup> entsprechen direkt der obigen Klassifikation der Ausfallprozesse und der Faktorisierung der gemeinsamen Wahrscheinlichkeitsverteilung für  $Y$ ,  $X$  und  $R$ :  $P(Y, X) P(R | Y, X)$ ; siehe Gleichung (3.1). Es werden zunächst Zusammenhänge und Verteilungen für die vollständigen Daten einschließlich der Ausfälle spezifiziert. Des Weiteren werden Annahmen über das Zustandekommen der Ausfälle formuliert. Hierbei können die Ausfälle sowohl von den beobachteten  $X$ -Werten als auch von den unbeobachteten  $Y$ -Werten abhängen.

Selektionsmodelle für metrische abhängige Variablen haben sich in der Praxis insbesondere aufgrund nicht überprüfbarer Verteilungsannahmen teilweise als problematisch herausgestellt (Little und Rubin 2002: 321-324; Stolzenberg und Relles 1997). Bei dem als Alternative dazu entwickelten Pattern-Mixture Modell wird die gemeinsame Wahrscheinlichkeit für  $Y$ ,  $X$  und  $R$  als  $P(Y, X | R) P(R)$  faktorisiert (Little und Rubin 2002: 312ff.; Molenberghs und Verbeke 2005: 555ff.). Aber nicht nur beim Selektionsmodell, sondern auch beim Pattern-Mixture Modell gehen nicht überprüfbare Annahmen und Restriktionen in die Schätzung ein, beispielsweise in der Weise, dass die Zusammenhänge zwischen abhängiger und erklärender Variablen in der Gruppe mit Ausfall denen der Gruppe ohne Ausfall entsprechen (Little 1993: 129). Als Vorteil von Pattern-Mixture Modellen wird gesehen, dass die Restriktionen offenkundiger sind als bei Selektionsmodellen. Ein weiterer Vorteil von Pattern-Mixture Modellen liegt darin, dass damit Unterschiede der Ausfallprozesse verschiedener Gruppen direkt untersucht werden können. Ist man jedoch an Aussagen über Zusammenhänge in der Gesamtheit (einschließlich Ausfällen) interessiert, ist das Selektionsmodell zu präferieren, da für die Modellierung der strukturellen Zusammenhänge das gleiche Modell wie im Falle ohne Ausfälle verwendet wird (Allison 2002: 79; Molenberghs und Verbeke 2005: 573).<sup>8</sup>

7 Der Begriff Selektionsmodell geht auf das Verfahren von Heckman (1976) zurück, das für fehlende oder unvollständig beobachtete metrische Variablen entwickelt wurde (Glynn et al. 1986; Molenberghs und Verbeke 2005: 484).

8 Mit Selektionsmodellen können ebenfalls Zusammenhänge für die Gruppen mit vollständigen Antworten vs. Ausfällen dargestellt werden, sodass sich in dieser Hinsicht Selektionsmodelle und Pattern-Mixture Modelle ähnlich sind (siehe Molenberghs et al. 1999: 113). Allerdings führen aufgrund unter-

Im Folgenden wird deshalb das der Klasse der Selektionsmodelle zugehörige log-lineare Modell bzw. die log-lineare Pfadanalyse verwendet,<sup>9</sup> mit der die Ausfallprozesse und strukturellen Zusammenhänge bei kategorialen Daten modellierbar sind. Die gemeinsame Wahrscheinlichkeit von  $Y$  und  $R$  in Abhängigkeit von  $X$  ist  $p_{yrlx}$ . Die Likelihood ist (Baker und Laird 1988: 63):

$$L = \left[ \prod_x \prod_y \left( p_{y1|x} \right)^{n_{xy1}} \right] \left[ \prod_x \left( p_{+0|x} \right)^{n_{x+0}} \right] \quad (3.5)$$

mit  $n_{xy1}$  Fallzahl der Merkmalskombination  $X$  und  $Y$  für die Gruppe ohne Ausfall ( $R = 1$ )  
 $n_{x+0}$  Fallzahl der Merkmalskombination  $X$  für die Ausgefallenen ( $R = 0$ ) mit unbekannten  $Y$ -Werten

Für ignorierbare Ausfälle (MCAR, MAR) entspricht der erste Teil der Likelihood einem Strukturmodell, in dem die substanzwissenschaftlich relevanten Variablen modelliert werden. Der zweite Teil der Likelihood kann als Ausfallmodell betrachtet werden. Die Voraussetzung für diese Faktorisierung der Likelihood besteht darin, dass der Ausfall  $R$  bedingt unabhängig von  $Y$  und  $X$  sein muss. Bei Paneldaten mit monotonem Ausfallmuster trifft dies zu (Baker und Laird 1988: 66). Als monotonen Ausfallmuster wird bezeichnet, wenn Einheiten im Zeitverlauf sukzessive ausfallen, ohne später wieder befragt zu werden und für sie vollständige Angaben zu den  $X$ -Variablen vorliegen. Auf Basis der vollständig beobachteten Daten können dann für ein gegebenes Muster von Ausfällen und Kovariaten die fehlenden  $Y$ -Werte geschätzt werden. Bei vollständig kategorialen Daten ist die Unterscheidung zwischen MAR- und NINR-Modellen allerdings nicht mehr bzw. nur eingeschränkt möglich, da unter MAR-Annahme die Koeffizienten des Ausfallmodells nicht mehr unabhängig von den Koeffizienten des Strukturmodells geschätzt werden können (Vermunt 1997a: 77).

Da  $Y$  für die Ausgefallenen ( $R = 0$ ) nicht beobachtbar ist, können für die Schätzung keine Standardverfahren der log-linearen Analyse benutzt werden. In den hier verwendeten speziellen log-linearen Modellen, bei denen die fehlenden Daten latente Klassen bzw. Zellenbesetzungen repräsentieren,<sup>10</sup> werden Schätzungen zumeist mithilfe des EM-Algorithmus (EM: Expectation, Maximisation) vorgenommen. Die Grundidee besteht darin, die Schätzung für unvollständige Daten durch eine einfachere iterative Schätzung für vollständige Daten zu ersetzen. In einem ersten E-Schritt werden die unter einem gegebenen Modell erwarteten Zellenbesetzungen für die Ausfallgruppe geschätzt. In einem zweiten M-Schritt erfolgt die Maximierung der Likelihood der vollständigen Daten (inkl. der Ausfälle). Der M-Schritt ergibt neue erwartete Zellenbesetzungen. Dies wird so lange wiederholt, bis sich die Schätzungen nicht mehr unterscheiden (Dempster et

---

schiedlicher Verteilungsannahmen Selektionsmodelle und Pattern-Mixture Modelle – obwohl sie mathematisch ineinander überführbar sind – mit Ausnahme vollständig zufälliger Ausfälle (MCAR) zu unterschiedlichen Modellen (Little 1993: 126; Toutenburg et al. 2004: 34-37). Während z. B. bei MAR-Annahme das Selektionsmodell  $f(X, Y, R) = f(X, Y)f(R|X)$  lautet, wird mit dem Pattern-Mixture Modell  $f(X, Y|R) = f(X, Y)f(R|X)/f(R)$  geschätzt (Toutenburg et al. 2004: 37).

- 9 Die Bezeichnung des Modells als log-lineares Pfadmodell bzw. Pfadanalyse folgt Vermunt (1997a). Sie geht auf das von Goodman (1973) entwickelte Modell zurück, in dem die gemeinsame Wahrscheinlichkeit von Variablen entsprechend von Überlegungen über ihre kausale Anordnung in ein Produkt von marginalen und bedingten Wahrscheinlichkeiten zerlegt wird.
- 10 Während in der üblichen Analyse latenter Klassen (LCA) jede Beobachtung zu einer potenziellen latenten Dimension gehört, stellen hier die fehlenden Werte latente Zellenbesetzungen dar (Winship et al. 2002: 413).



al. 1977; Little und Rubin 2002: 164ff.; Schafer 1997: 37ff.); siehe dazu auch das Beispiel auf Seite 31. Im E-Schritt der Schätzung des obigen Modells (Gl. 3.5) wird die folgende Reparametrisierung verwendet (Baker und Laird 1988):

$$\begin{aligned} p_{y|x} &= p_{y|x} p_{r|yx} \\ p_{y|x} &= m_{xy} / m_{x+} \\ p_{r|yx} &= M_{xy} / M_{xy+} \end{aligned} \quad (3.6)$$

Mit  $m$  werden die erwarteten Zellenbesetzungen für die Teiltabelle  $R = 0$  und mit  $M$  die erwarteten Zellenbesetzungen der Gesamttabelle bezeichnet.

Zur Lösung der Schätzprobleme bei unvollständigen Daten kann neben dem EM-Algorithmus und bayesianischen Methoden (siehe Little und Rubin 2002; Park und Brown 1994; Forster und Smith 1998) auch ein sogenanntes „Composite Link-Model“ eingesetzt werden (siehe Baker 1994; Chambers und Welsh 1993; Rindskopf 1992, Molenberghs und Goetghebeur 1997). Mit diesem log-linearen Modell werden in einer speziellen Matrix Restriktionen gesetzt, mit denen die linearen Beziehungen zwischen den vollständigen und den unvollständigen Daten abgebildet werden (siehe als einfaches Beispiel für die Behandlung der linearen Beziehungen Seite 34). Die Konstruktion dieser Restriktionsmatrizen kann jedoch sehr aufwendig sein.

### 3.3 Beispiele

Das gewählte log-lineare Modell wird im Folgenden anhand eines einfachen Beispiels von Daten des Mikrozensuspanels 1996-1999 aus Kapitel 6, Abschnitt 6.1, beschrieben. Tabelle 3.1 berichtet für Auszubildende des dualen Systems im April 1996 deren Ausbildungsstatus ( $Y$ ) im April 1997 in Abhängigkeit von der Beschäftigungsdauer ( $X$ ). Von insgesamt 1.381 Auszubildenden liegen für 1.123 (81,3 %) räumlich immobile Personen ohne Ausfall ( $R = 1$ ) Angaben zu ihrem Ausbildungsstatus im April 1997 vor. Für 258 Auszubildende (18,7 %), die zwischen den Mikrozensususerhebungen 1996 und 1997 fortgezogen sind ( $R = 0$ ), fehlen diese Informationen.



Tabelle 3.1: Ausbildungsstatus im April 1997 nach Beschäftigungsdauer für Auszubildende im April 1996 im Mikrozensuspanel 1996-1999 – Fallzahlen und Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	mit Angaben zum Ausbildungsstatus (Y)				ohne Angabe	Ins- gesamt
	1 Auszub.	2 Abschluss	3 Abbruch	Insges. (R = 1)	Insges. (R = 0)	(R = 0, 1)
1 ≤12 Monate	420 (85,2)	19 (3,9)	54 (11,0)	493 (100)	92	585
2 13-24 Monate	282 (67,3)	94 (22,4)	43 (10,3)	419 (100)	85	504
3 25+ Monate	25 (11,8)	167 (79,1)	19 (9,0)	211 (100)	81	292
Insgesamt	727 (64,7)	280 (24,9)	116 (10,3)	1.123 (100)	258	1.381

Quelle: Mikrozensuspanel 1996-1999; eigene Berechnungen (siehe Abschnitt 6.1).

Mit der Ausnahme von MCAR-Modellen können die Schätzungen für diese quadratische (3 x 3 -) Tabelle der Gesamtheit in geschlossener Form dargestellt werden (Baker et al. 1992; Little und Rubin 2002: 268f., 344).<sup>11</sup> Damit sind die Schätzungen, die mit dem verwendeten Programm LEM (Vermunt 1997b) mittels EM-Algorithmus durchgeführt werden, anschaulich nachvollziehbar. Die LEM-Programme zu den in diesem Kapitel geschätzten Modellen sind im Anhang wiedergegeben.

Es wird folgende Notation verwendet. Die entsprechenden Laufindizes der Variablen  $X$ ,  $Y$  und  $R$  sind  $i$ ,  $j$  und  $k$ . Die beobachteten bzw. geschätzten Zellenbesetzungen sind dann  $n_{ijk}$  bzw.  $m_{ijk}$ . Das Zeichen „+“ steht für die Summierung über dem entsprechenden Index.

Wie die Teiltabelle (3.1) der Befragten mit vollständigen Angaben ( $R = 1$ ) zeigt, hängen die Beschäftigungsdauer und der Ausbildungsstatus eng zusammen. Beträgt die Beschäftigungsdauer im Jahre 1996 bis zu zwölf Monate, befinden sich 1997 noch rund 85 Prozent ( $420 / 493 = 0,85$ ) der Auszubildenden in der Ausbildung. Dagegen haben 1997 rund 80 Prozent ( $167 / 211 = 0,79$ ) der Personen mit einer bisherigen Beschäftigungsdauer von wenigstens 25 Monaten den Abschluss erreicht. Zwischen 1996 und 1997 brechen in jeder Klasse der Beschäftigungsdauer rund zehn Prozent die Ausbildung ab.

Wird angenommen, dass der erreichte Ausbildungsstatus ( $Y$ ) von der Beschäftigungsdauer ( $X$ ) abhängt, entspricht das Strukturmodell dem log-linearen Modell  $\log(m_{ij}) = \alpha + \beta_i^X + \beta_j^Y + \beta_{ij}^{XY}$ , kurz  $\{XY\}$ , bzw. dem äquivalenten Logit-Modell für  $P(Y = j | X) = m_{ij} / \sum_j m_{ij}$ .

Das Ausfallmodell für  $P(R=0)$  ist ebenfalls ein log-lineares bzw. Logit-Modell, das je nach Annahmen zum Ausfallzusammenhang Effekte für  $X$  bzw.  $Y$  enthalten kann. Wegen der Konzen-

11 Dass gerade bei dem einfachsten Ausfallmodell MCAR keine Schätzung in geschlossener Form durchführbar ist, erscheint überraschend, liegt jedoch an den Restriktionen log-linearer Modelle der Übereinstimmung von beobachteten und geschätzten Randverteilungen der Gesamttabelle. Z. B. entsprechen aufgrund der unterschiedlichen Randverteilungen der  $X$ -Variablen der Gruppen ohne und mit Ausfall bei einfacher Berechnung der Zellenbesetzungen der Ausfallgruppe durch  $m_{j0} = (m_{ijl} / m_{i+1}) * m_{++0}$  die geschätzten Randverteilungen der  $X$ -Variablen für die Gesamttabelle nicht den beobachteten Werten. Berücksichtigt man die unterschiedliche Verteilung der  $X$ -Variablen in den Teiltabellen z. B. durch die Berechnung mit  $m_{j0} = (m_{ijl} / m_{i+1}) * m_{i+0}$ , unterscheiden sich die Randverteilungen der  $Y$ -Variablen in den Gruppen ohne und mit Ausfall, was der MCAR-Annahme  $P(Y=j | R=0) = P(Y=j | R=1)$  widerspricht.

tration auf eine explizite abhängige Variable wird das log-lineare Modell als Logit-Modell spezifiziert und die Notation des LEM-Programms verwendet:  $Y|X \{XY\} R|XY \{R|XY\}$ . Der erste Ausdruck  $Y|X \{XY\}$  beschreibt das Strukturmodell und der zweite Ausdruck  $R|XY \{R|XY\}$  das Ausfallmodell.

### 3.3.1 Vollständig zufälliger Ausfall (MCAR)

Unter vollkommen zufälligen Ausfällen (MCAR), d. h. wenn in diesem Beispiel Wegzüge weder mit der Beschäftigungsdauer noch mit dem Ausbildungsstatus zusammenhängen, gilt  $P(R|X, Y) = P(R)$ . Das entsprechende Logit-Modell lautet:  $Y|X \{XY\} R|X\{R\}$ . Somit wird für alle Merkmalskombinationen ein konstantes Verhältnis von Ausfällen ( $R = 0$ ) und vollständigen Angaben ( $R = 1$ ) angenommen:  $b = P(R = 0) / P(R = 1) = m_{ij0} / m_{ij1}$ . In diesem Fall beträgt  $b = 0,2297$  ( $= 258 / 1.123$ ; siehe Tab. 3.1).

Die Zellenbesetzungen  $m_{ij}$  und Anteile  $\theta_{ij}$  werden mit einem einfachen EM-Verfahren geschätzt. Im E-Schritt werden die unter dem MCAR-Modell erwarteten Zellenbesetzungen  $m_{ij+}$  der Gesamttabelle wie folgt geschätzt:

$$\hat{E}(m_{ij+} | Y_{obs}, \theta) = \hat{E}(m_{ij1} + m_{ij0} | Y_{obs}, \theta) = n_{ij1} + n_{i+0} \hat{\theta}_{ij} / \hat{\theta}_{i+}$$

Im M-Schritt, mit dem die Likelihood der vervollständigten Daten maximiert wird, erfolgt die Schätzung der Anteile  $\theta_{ij}$  durch:

$$\hat{\theta}_{ij} = \hat{E}(m_{ij+} | Y_{obs}, \theta) / n_{+++}$$

Kombiniert man den E- und M-Schritt ergibt sich für eine Iteration  $t$  (Schafer 1997: 44):

$$\hat{\theta}_{ij}^{(t+1)} = \left[ n_{ij1} + n_{i+0} \left( \hat{\theta}_{ij}^t / \hat{\theta}_{i+}^t \right) \right] / n_{+++}$$

Mit den Zellenbesetzungen bzw. Anteilen der räumlich Immobilen ( $R = 1$ ) als Startwerten erhält man nach zehn Iterationen das Endergebnis.

Alternativ kann in diesem Fall der quadratischen Tabelle auch das Verfahren von Baker et al. (1992: 647 [Model (a)]) angewendet werden. Der kombinierte E- und M-Schritt für die Schätzung der Zellenbesetzungen der räumlich Immobilen  $m_{ij1}$  mit

$$\hat{m}_{ij1}^{(t+1)} = \left( (n_{ij1} + n_{i+0} \hat{m}_{ij1}^t / \hat{m}_{i+1}^t) n_{++1} \right) / n_{+++}$$

und  $m_{ij1}^{(t=1)} = n_{ij1}$  als Startwerten ergibt bereits bei der zweiten Iteration das Endergebnis. Durch Multiplikation  $m_{ij0} = b * m_{ij1}$  erhält man die Fallzahlen für die Ausfallgruppe ( $R = 0$ ).

Tabelle 3.2 zeigt hierfür die nach der Beschäftigungsdauer ( $X$ ) bedingten Anteile des Ausbildungsstatus ( $Y$ ), die modellgemäß für beide Teilgruppen identisch sind. Wie aber der Vergleich mit Tabelle 3.1 zeigt, weichen bei diesem MCAR-Modell die Randverteilungen der geschätzten von denen der beobachteten Daten in den Teilgruppen ab. Aus Platzgründen werden keine Regressionskoeffizienten, sondern nur die geschätzten Anteilswerte und Randverteilungen sowie Kennziffern der Modellschätzung berichtet.

Als Maß der Modellanpassung wird die Devianz  $G^2 = -2 * (l_M - l_{beob})$  verwendet, mit der die Log-Likelihood eines gegebenen Modells  $M$  mit der Log-Likelihood eines saturierten Modells verglichen wird, das Gleichheit von erwarteten und beobachteten Zellenbesetzungen (*beob*) un-

terstellt. Da sich die statistische Inferenz auf die beobachteten Verteilungen beziehen soll, werden bei der Berechnung der Devianz die einzelnen geschätzten  $Y$ -Werte der Ausfallgruppe zu einem Wert zusammengefasst. Bei kleinen Zellenbesetzungen kann die übliche Annahme einer Chi-Quadrat-Verteilung verletzt sein. Weniger trifft dies auf die Likelihood-Ratio-Teststatistik  $G^2 = -2 \cdot (l_{M1} - l_{M2})$  zu, mit der die Differenz der Devianzen verschiedener Modelle berechnet wird. Die Datentabelle 3.1 umfasst zwölf Zellen. Mit dem MCAR-Modell  $Y|X \{XY\} R|XY \{R\}$  werden zehn log-lineare Parameter geschätzt ( $\alpha, \beta_2^X, \beta_3^X, \beta_2^Y, \beta_3^Y, \beta_{22}^{XY}, \beta_{23}^{XY}, \beta_{32}^{XY}, \beta_{33}^{XY}, \beta_0^R$ ). Somit verbleiben zwei Freiheitsgrade ( $d.f.$ ; siehe Tabelle 3.2).

Tabelle 3.2: Unter der Annahme vollständig zufälliger Ausfälle (MCAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			
	1 Auszub.	2 Abschluss	3 Abbruch	Insgesamt
mit Angaben zum Ausbildungsstatus (R = 1)				
1 ≤12 Monate	85,2	3,9	11,0	475,7
2 13-24 Monate	67,3	22,4	10,3	409,8
3 25+ Monate	11,8	79,1	9,0	237,4
Insgesamt	63,2	26,6	10,3	1.123
ohne Angaben zum Ausbildungsstatus (R = 0)				
1 ≤12 Monate	85,2	3,9	11,0	109,3
2 13-24 Monate	67,3	22,4	10,3	94,2
3 25+ Monate	11,8	79,1	9,0	54,6
Insgesamt	63,2	26,6	10,3	258
Insgesamt (R = 1,0)	63,2	26,6	10,3	1.381

Datenbasis: Tabelle 3.1. Modellspezifikation (Logit-Modell):  $Y|X \{XY\} R|XY \{R\}$ ; Devianz  $G^2 = 18,84$ ;  $d.f. = 2$ ; Log-Likelihood = -2.200,99; Anzahl log-linearer Parameter = 10.

3.3.2 Bedingt zufälliger Ausfall (MAR)

Es ist zu vermuten, dass die Ausfälle infolge räumlicher Mobilität nach erfolgreichem Abschluss, d. h. bei längerer Beschäftigungsdauer, größer sind als für Auszubildende, die noch am Anfang der betrieblichen Ausbildung stehen. Im nächsten Modell soll deshalb ein bedingt zufälliger Ausfall (MAR) angenommen werden, bei dem  $P(Y = j | X = i, R = 0) = P(Y = j | X = i, R = 1)$  gilt. Das Logit-Modell  $Y|X \{XY\} R|XY \{RX\}$  enthält für die Schätzung dieser Zusammenhänge die Interaktion zwischen  $R$  und  $X$ .

Für diese Beispieldaten liefert bereits die einfache getrennte Schätzung von Struktur- und Ausfallmodell (Gl. 3.5) alle Informationen, um die bedingten Wahrscheinlichkeiten der Teiltabelle ohne Ausfälle  $P(Y | X, R = 1)$  auf die Teiltabelle mit Ausfällen ( $R = 0$ ) entsprechend der pro  $X$ -Kategorie geschätzten Ausfallwahrscheinlichkeit  $P(R = 0 | X)$  und der Verteilung von  $X(m_{i+0})$  in dieser Teiltabelle zu übertragen (Little und Rubin 2002: 266ff.). Die Zellenbesetzungen für die Ausfallgruppe sind einfach wie folgt zu schätzen:  $m_{ij0} = \hat{p}_{j|i,R=1} m_{i+0}$ , z. B.  $m_{110} = (420/493) \cdot 92 = 78,4$ .

Die Ausfallwahrscheinlichkeit beträgt für eine Beschäftigungsdauer von bis zu zwei Jahren 16 bzw. 17 Prozent, bei 25 und mehr Monaten jedoch rund 28 Prozent ( $81 / (81 + 211) = 0,277$ ; siehe Tab. 3.1). Gegenüber dem MCAR-Modell verbraucht das MAR-Modell zwei Parameter mehr, die

Likelihood-Ratio-Statistik zeigt aber mit  $G^2 = 18,84$  ( $= 2 \cdot (-2.191,57 - -2.200,99)$ ) eine signifikant bessere Modellanpassung an, sodass die MCAR-Annahme einer konstanten Ausfallwahrscheinlichkeit abzulehnen ist.

Tabelle 3.3: Unter der Annahme bedingt zufälliger Ausfälle (MAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			Insgesamt
	1 Auszub.	2 Abschluss	3 Abbruch	
mit Angaben zum Ausbildungsstatus (R = 1)				
1 ≤12 Monate	85,2	3,9	11,0	493
2 13-24 Monate	67,3	22,4	10,3	419
3 25+ Monate	11,8	79,1	9,0	211
Insgesamt	64,7	24,9	10,3	1.123
ohne Angaben zum Ausbildungsstatus (R = 0)				
1 ≤12 Monate	85,2	3,9	11,0	92
2 13-24 Monate	67,3	22,4	10,3	85
3 25+ Monate	11,8	79,1	9,0	81
Insgesamt	56,3	33,6	10,1	258
Insgesamt (R = 1,0)	63,2	26,6	10,3	1.381

Datenbasis: Tabelle 3.1. Modellspezifikation (Logit-Modell):  $Y|X \{XY\} R|XY \{RX\}$ ;

Devianz  $G^2 = 0$ ; d.f. = 0; Log-Likelihood = -2.191,57; Anzahl log-linearer Parameter = 12.

### 3.3.3 Nicht ignorierbarer Ausfall (NINR)

Wenn Jugendliche am Ende der Ausbildung, d. h. nach etwa drei Jahren und nach erfolgreichem Abschluss i. d. R. über ein höheres Einkommen verfügen können, dürften sie gegenüber Auszubildenden eine höhere Umzugsneigung aufweisen. Ähnlich dürfte die Umzugsneigung von Ausbildungsabbrechern steigen, sofern sie den bisherigen Ausbildungsbetrieb verlassen. Es liegt deshalb nahe, einen Zusammenhang zwischen dem Ausfall bzw. Wegzug und dem Ausbildungsstatus und damit nicht ignorierbare Ausfälle (NINR bzw. MNAR) anzunehmen.

Ist nun die Ausfallwahrscheinlichkeit nicht für alle  $Y$ -Werte konstant, sondern hängt von den fehlenden Werten selbst ab  $P(R | Y_{obs}, Y_{mis}) \neq P(R | Y_{obs})$ , wird bei der Schätzung des NINR-Modells  $Y|X \{XY\} R|XY \{RY\}$  davon ausgegangen, dass die geschätzten Spaltenprozentwerte  $m_{ij}/m_{+j}$  für die Gruppen ohne und mit Ausfall gleich sind (Little und Rubin 2002: 344):

$$m_{ij0} / m_{+j0} = m_{ij1} / m_{+j1}$$

Zusammen mit der Bedingung, dass bei der Ausfallgruppe die geschätzten Randverteilungen mit den beobachteten Werten übereinstimmen müssen:

$$\sum_j m_{ij0} = m_{i0}$$

(siehe Gleichungen 7-9 unten), ergibt sich, dass die Verhältnisse (Odds) der geschätzten Zellenbesetzungen für die Ausfallgruppe ( $m_{ij0}$ ), den Odds der Zellenbesetzungen der vollständig beob-

achteten Tabelle ( $m_{ijl}$ ) entsprechen sollen (Gleichungen 1-6). Dies veranschaulicht, dass damit die Verteilungen von  $Y$  der vollständig beobachteten Daten ( $Y_{obs}$ ) auf die Daten der Ausfälle ( $Y_{mis}$ ) übertragen werden:

$$(R\ 1) \quad m_{210}/m_{110} = m_{211}/m_{111}$$

$$(R\ 2) \quad m_{310}/m_{110} = m_{311}/m_{111}$$

$$(R\ 3) \quad m_{220}/m_{120} = m_{221}/m_{121}$$

$$(R\ 4) \quad m_{320}/m_{120} = m_{321}/m_{121}$$

$$(R\ 5) \quad m_{230}/m_{130} = m_{231}/m_{131}$$

$$(R\ 6) \quad m_{330}/m_{130} = m_{331}/m_{131}$$

$$(R\ 7) \quad m_{110} + m_{120} + m_{130} = n_{1+0}$$

$$(R\ 8) \quad m_{210} + m_{220} + m_{230} = n_{2+0}$$

$$(R\ 9) \quad m_{310} + m_{320} + m_{330} = n_{3+0}$$

Da in diesem Beispiel für die Gruppe ohne Ausfälle im Strukturmodell ein saturiertes Modell  $\{XY\}$  geschätzt wird, entsprechen die geschätzten Fallzahlen den beobachteten Werten. Die obigen Restriktionen (R1-R9) bilden die Matrix **A**.<sup>12</sup>

$$\begin{pmatrix} 0,6714 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0,0595 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 4,9474 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 8,7895 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0,7963 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0,3519 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} m_{110} \\ m_{120} \\ m_{130} \\ m_{210} \\ m_{220} \\ m_{230} \\ m_{310} \\ m_{320} \\ m_{330} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 92 \\ 85 \\ 81 \end{pmatrix}$$

**A m = n**

Das in der ersten Restriktion R1 relevante Verhältnis der Zellenbesetzungen der vollständigen Tabelle beträgt  $n_{211}/n_{111} = 282/420 = 0,6714$ . Damit lässt sich R1 als  $m_{210} = 0,6714 * m_{110}$  formulieren. Eine weitere Umformung entspricht der ersten Zeile des obigen linearen Gleichungssystems:  $0,6714 * m_{110} - m_{210} = 0$ . Analog dazu werden die anderen Restriktionen bis R6 dargestellt. Die restlichen Zeilen der Matrix **A** entsprechen den Restriktionen R7 bis R9. Löst man dieses Gleichungssystem mit  $\mathbf{m} = \mathbf{A}^{-1} \mathbf{n}$ , erhält man unter bestimmten Bedingungen die gesuchten Zellenbesetzungen.

Mit dieser Beispieltabelle, in der die Odds der beobachteten Randverteilung ausgefallener Daten nahezu gleich sind ( $m_{210}/m_{110} = 0,92$  und  $m_{310}/m_{110} = 0,88$ ), werden allerdings, wie es häufig bei Datenausfall vorkommen kann, negative Zellenbesetzungen geschätzt. Dies liegt daran, dass die Odds der Randverteilung nicht im Intervall der Odds beobachteter Daten liegen (z. B.  $m_{211}/m_{111} = 0,67$  und  $m_{311}/m_{111} = 0,06$ ). Dann müssen eine oder mehrere Zellenbesetzungen auf Null gesetzt werden und es liegt eine sogenannte „boundary solution“ oder Stützung vor (Baker et al. 1992; Little und Rubin 2002: 344).

12 Für diesen Hinweis danke ich Siegfried Gabler.

Mithilfe des EM-Algorithmus können allerdings in diesen Fällen immer Lösungen gefunden werden, deren Werte größer als Null sind, also im zulässigen Wertebereich liegen (Baker und Laird 1988). Tabelle 3.4 enthält die entsprechenden Ergebnisse des LEM-Programms.

Für Absolventen wird durch das NINR-Modell die Ausfallwahrscheinlichkeit  $P(R = 0 | Y = j)$  auf 23 Prozent), für Ausbildungsabbrecher auf 60 Prozent geschätzt.<sup>13</sup> Von der Ausfallgruppe werden dementsprechend die meisten Personen den Abbrechern zugeordnet. Bei den vollständig beobachteten Daten weichen aufgrund der Schätzprobleme die geschätzten von den beobachteten Odds etwas ab, weshalb die Devianz  $G^2$  geringfügig von Null verschieden ist. Um Schätzungen der Gesamttabelle inkl. der systematischen Ausfälle zu erhalten, können die Daten der beobachteten Werte  $n_{ijl}$  auch mit dem Kehrwert der durch dieses Modell geschätzten Antwortwahrscheinlichkeiten  $(1 - P(R = 0 | Y = j))$  gewichtet werden.

Für Auszubildende wird durch das NINR-Modell die Ausfallwahrscheinlichkeit auf nahezu Null Prozent geschätzt (siehe Tabelle 3.4:  $(0,1\% * 258) / (52,7\% * 1.381)$ ). Geht man davon aus, dass Jugendliche auch während der Ausbildung umziehen, erscheint dieses Teilergebnis wenig plausibel. Es liegt deshalb nahe, zusätzlich Zusammenhänge zwischen dem Ausfall und der Beschäftigungsdauer zu schätzen. Weil aber sowohl das obige MAR-Modell als auch das NINR-Modell bereits alle Freiheitsgrade verbraucht haben, kann der nicht ignorierbare Ausfall für diese Beispieldaten nicht zusätzlich zur MAR-Annahme modelliert werden. Stehen aber weitere erklärende Variablen zur Verfügung, können Freiheitsgrade durch Restriktionen bzw. den Ausschluss höherer Interaktionen gewonnen werden.

Tabelle 3.4: Unter der Annahme nicht ignorierbarer Ausfälle (NINR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			Insgesamt
	1 Auszub.	2 Abschluss	3 Abbruch	
mit Angaben zum Ausbildungsstatus (R = 1)				
1 ≤12 Monate	84,8	3,9	11,3	495
2 13-24 Monate	67,9	22,4	9,8	415
3 25+ Monate	11,8	79,0	9,2	213
Insgesamt	64,7	24,9	10,3	1.123
ohne Angaben zum Ausbildungsstatus (R = 0)				
1 ≤12 Monate	0,1	6,3	93,5	90
2 13-24 Monate	0,1	31,1	68,8	89
3 25+ Monate	0,0	62,9	37,1	79
Insgesamt	0,1	32,2	67,7	258
Insgesamt (R = 1,0)	52,7	26,3	21,0	1.381

Datenbasis: Tabelle 3.1. Modellspezifikation (Logit-Modell):  $Y|X \{XY\} R|XY \{RY\}$ ; Devianz  $G^2 = 0,48$ ; d.f. = 0; Log-Likelihood = -2.191,81; Anzahl log-linearer Parameter = 12; „boundary solution“.

Bei jeweils null Freiheitsgraden sind die Devianzen der MAR- und NINR-Modelle nahezu gleich (siehe Tab. 3.3 und 3.4). Wie häufig bei Daten mit fehlenden Werten (Little und Rubin 2002: 344),

13 Diese Anteile werden vom LEM-Programm berichtet. Sie können auch anhand der Angaben in Tabelle 3.4 berechnet werden; z. B. für die Absolventen:  $(32,2\% * 258) / (26,3\% * 1.381) = 23$  Prozent.

liegen auch in diesem Beispiel für eine Entscheidung zwischen den verschiedenen Modellannahmen praktisch keine statistischen Kriterien vor. Man kann nun die Schätzergebnisse auf ihre Plausibilität hinterfragen.

In dem obigen NINR-Modell z. B. wird für Auszubildende eine Ausfallwahrscheinlichkeit von fast null Prozent geschätzt. Dies erscheint wenig plausibel. Für solche Überlegungen sind Sensitivitätsanalysen hilfreich, mit denen gezeigt werden kann, in welchem Bereich die Ergebnisse von verschiedenen für plausibel gehaltenen Modellen variieren (Allison 2002; Little und Rubin 2002: 344; Molenberghs und Verbeke 2005: 575ff.). Doch letztlich kann über die Gültigkeit der Ergebnisse nur spekuliert werden, falls keine externen Daten als Prüfstandard („gold standard“) verfügbar sind.

### 3.3.4 Modelle mit externen Referenzdaten

Es gibt verschiedene Vorschläge, wie externe Referenzdaten bei Ausfallanalysen berücksichtigt werden können. Z. B. verwenden Fitzmaurice et al. (1996) Querschnittsdaten bei der Analyse von Panelausfällen zur Verbesserung der Identifikation von Koeffizienten zu nicht ignorierbaren Ausfällen bei der Analyse von Wahlabsichten. Querschnittsinformationen des Mikrozensus wurden ebenfalls im Verbundprojekt zum Mikrozensuspanel zur Validierung eingesetzt. Darüber hinaus wurde für die Validierung u. a. das Sozioökonomische Panel (SOEP) genutzt (Basic und Rendtel 2005: 8ff.). Durch den Vergleich räumlich immobiler Personen mit den gesamten SOEP-Daten ergaben sich einerseits Hinweise auf den im Mikrozensuspanel bei der Analyse räumlich Immobiler zu erwartenden Bias. Andererseits lieferten die mit den SOEP-Daten beobachteten Verläufe Restriktionen, die in log-linearen Modellen zum Ausfallgeschehen im Mikrozensuspanel zur Identifikation von Parametern verwendet werden konnten (Basic et al. 2005). Dabei müssen allerdings Abweichungen aufgrund unterschiedlicher Stichprobenpläne sowie evtl. Befragungsausfälle im SOEP hingenommen werden.

In dieser Arbeit werden SOEP-Daten nicht verwendet, da damit einerseits Besucher der gymnasialen Oberstufe nicht äquivalent zum Mikrozensus abgegrenzt werden können. Andererseits wären die Fallzahlen des SOEP für Analysen zu beruflichen Ausbildungsverläufen zu klein. Als Referenzdaten dienen Querschnittsangaben der Bildungsstatistik bzw. die Beschäftigtenstichprobe (siehe Kapitel 2, Abschnitte 2 und 3).

In der Beschäftigtenstichprobe liegt zwar im Unterschied zum SOEP keine Information über die räumliche Mobilität von Auszubildenden vor, jedoch können die Daten als Stichprobe der Gesamtheit betrachtet werden, die nicht durch die bei freiwilligen Umfragen auftretenden Probleme von Befragungsausfällen beeinträchtigt ist. Wie die Verlaufsangaben der Beschäftigtenstichprobe und des Mikrozensuspanels in log-linearen Modellen zur Untersuchung selektiver Ausfälle eingesetzt werden können, wird für die Beispieldaten im Folgenden skizziert. Die Konstruktion der hier verwendeten Daten aus der Beschäftigtenstichprobe ist in Abschnitt 6.1 beschrieben.

Vergleicht man die Verteilungen in der Beschäftigtenstichprobe mit denen des Mikrozensuspanels (ohne Ausfälle), ist die Beschäftigungsdauer näherungsweise gleich verteilt. In der Beschäftigtenstichprobe finden sich jedoch etwas weniger Ausbildungsabsolventen (5.095 / 23.070 = 22 %) und mehr Abbrecher (14 %) als im Mikrozensuspanel (25 % bzw. 10 %; siehe Tab. 3.5). Stärkere Abweichungen sind bei einer Beschäftigungsdauer von wenigstens 25 Monaten zu beobachten. Während in der Beschäftigtenstichprobe für diese Gruppe die Anteile 61 Prozent für Abschluss und 26 Prozent für Abbruch betragen, liegen sie im Mikrozensuspanel bei 79 bzw. neun Prozent. Es ist zu vermuten, dass diese Differenzen mit dem Ausfall zusammenhängen.



Der Indikator für den Datentyp ( $D$ : 1 = Beschäftigtenstichprobe, 2 = Mikrozensuspanel) kann beim Vergleich wie eine unabhängige Variable behandelt werden. Bei der Modellierung sind zwei mögliche Vorgehensweisen zu unterscheiden.

Analog zu den obigen Beispielen kann man lediglich das Insgesamt ohne Berücksichtigung des Datentyps verwenden. Allerdings wird dabei nicht berücksichtigt, dass die hier verwendeten Daten der Beschäftigtenstichprobe aufgrund der Recodierungen (per Definition) keine Ausfälle enthalten (siehe S. 120 und Fußnote 60). Um dies zu berücksichtigen, sind Ausfälle in der Beschäftigtenstichprobe als strukturelle Nullzellen zu behandeln. Das lässt sich durch die Interaktion  $RD$  formulieren.

*Tabelle 3.5:* Ausbildungsstatus im April 1997 nach Beschäftigungsdauer für Auszubildende im April 1996 im Mikrozensuspanel 1996-1999 und in der Beschäftigtenstichprobe – Fallzahlen und Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	mit Angaben zum Ausbildungsstatus (Y)			ohne Ang.		
	1 Auszub.	2 Abschluss	3 Abbruch	Insges. (R = 1)	Insges. (R = 0)	Insgesamt
Beschäftigtenstichprobe (D = 1)						
1 ≤ 12 Monate	8.710 (87,5)	399 (4,0)	841 (8,5)	9.950 (100)		9.950
2 13-24 Monate	5.453 (70,7)	1.377 (17,9)	876 (11,4)	7.706 (100)		7.706
3 25+ Monate	668 (12,3)	3.319 (61,3)	1.427 (26,4)	5.414 (100)		5.414
Insgesamt	14.831 (64,3)	5.095 (22,1)	3.144 (13,6)	23.070 (100)		23.070
Mikrozensuspanel (D = 2)						
1 ≤ 12 Monate	420 (85,2)	19 (3,9)	54 (11,0)	493 (100)	92	585
2 13-24 Monate	282 (67,3)	94 (22,4)	43 (10,3)	419 (100)	85	504
3 25+ Monate	25 (11,8)	167 (79,1)	19 (9,0)	211 (100)	81	292
Insgesamt	727 (64,7)	280 (24,9)	116 (10,3)	1.123 (100)	258	1.381
Insgesamt						
1 ≤ 12 Monate	9.130	418	895	10.443	92	10.535
2 13-24 Monate	5.735	1.471	919	8.125	85	8.210
3 25+ Monate	693	3.486	1.446	5.625	81	5.706
Insgesamt	15.558	5.375	3.260	24.193	258	24.451

*Quelle:* Beschäftigtenstichprobe (IABS-R01) und Mikrozensuspanel 1996-1999; eigene Berechnungen (siehe Abschnitt 6.1).



Dementsprechend sind dann die Modelle bei den verschiedenen Annahmen zum Ausfallzusammenhang wie folgt zu spezifizieren:

Vollständig zufälliger Ausfall (MCAR):	$Y DX \{XY\} R DXY \{RD\}$
Bedingt zufälliger Ausfall (MAR):	$Y DX \{XY\} R DXY \{RDX\}$
Nicht ignorierbarer Ausfall (NINR):	$Y DX \{XY\} R DXY \{RDY\}$

Aufgrund des erheblich größeren Stichprobenumfangs der Beschäftigtenstichprobe werden die Schätzungen des Strukturmodells im Wesentlichen durch die Zusammenhänge in der Referenzstatistik geprägt sein. Mit dem Teilmodell  $Y|DX \{XY\}$  wird angenommen, dass die strukturellen Zusammenhänge zwischen  $Y$  und  $X$  in beiden Stichproben gleich sind. Da die Daten aber auf ganz unterschiedliche Weise entstanden sind (Umfragedaten vs. prozessproduzierte Daten) ist dies eine kritische Annahme. In Bezug auf die Konstruktion vergleichbarer Variablen und Abgrenzungen in beiden Stichproben stellen sich deshalb hohe Anforderungen und es kann grundsätzlich nicht ausgeschlossen werden, dass sich die strukturellen Zusammenhänge zwischen dem Status ( $Y$ ) und den erklärenden Variablen unterscheiden. Dennoch bietet diese Verwendung externer Referenzdaten gegenüber Sensitivitätsanalysen lediglich auf Basis des Mikrozensuspanels entscheidende Vorteile, da Sensitivitätsanalysen nur einen Eindruck über die Variabilität der Schätzergebnisse unter verschiedenen möglichen Ausfallannahmen geben können.

Für den Teil des Strukturmodells gilt zwar die obige Annahme der Verteilungsgleichheit, doch werden im MAR-Ausfallmodell aufgrund der Interaktion  $RDX$  eventuelle Verteilungsabweichungen der erklärenden Variablen in beiden Daten durch das spezifizierte Logit-Modell kontrolliert, da alle Interaktionen im Modell implizit enthalten sind.

Tabelle 3.6 zeigt die Ergebnisse, die unter der Annahme bedingt zufälliger Ausfälle im Mikrozensuspanel geschätzt werden, wenn im Strukturmodell zusätzlich die Daten der Beschäftigtenstichprobe enthalten sind.

Gegenüber einem Modell mit der Annahme vollständig zufälliger Ausfälle (MCAR;  $G^2 = 68,89$ ; d.f. = 8; ohne Darstellung) wird durch das MAR-Modell  $Y|DX \{XY\} R|DXY \{RDX\}$  mit  $G^2 = 50,04$  (d.f. = 6) eine statistisch signifikant bessere Anpassung an die beobachteten Verteilungen erzielt. Im Vergleich zum MAR-Modell, das nur die Daten des Mikrozensuspanels enthält (siehe Tab. 3.3), zeigen sich insbesondere bei den Anteilen für die Absolventen und Ausbildungsabbrecher mit der höchsten Beschäftigungsdauer deutliche Abweichungen in Höhe von jeweils rund 17 Prozentpunkten, die auf die Berücksichtigung der Referenzdaten zurückzuführen sind. Während hier von den ausgefallenen Personen mit einer Beschäftigungsdauer von wenigstens 25 Monaten 62 Prozent den Absolventen und 26 Prozent den Abbrechern zugeordnet werden, sind es in Tabelle 3.3 79 bzw. neun Prozent.

Tabelle 3.6: Unter der Annahme bedingt zufälliger Ausfälle (MAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			
	1 Auszub.	2 Abschluss	3 Abbruch	Insgesamt
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Beschäftigtenstichprobe ( $D=1$ )				
1 ≤12 Monate	87,4	4,0	8,6	9.950
2 13-24 Monate	70,6	18,1	11,3	7.706
3 25+ Monate	12,3	62,0	25,7	5.414
Insgesamt	64,2	22,3	13,5	23.070
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤12 Monate	87,4	4,0	8,6	493
2 13-24 Monate	70,6	18,1	11,3	419
3 25+ Monate	12,3	62,0	25,7	211
Insgesamt ( $R = 1$ )	67,0	20,2	12,8	1.123
ohne Angaben zum Ausbildungsstatus ( $R = 0$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤12 Monate	87,4	4,0	8,6	92
2 13-24 Monate	70,6	18,1	11,3	85
3 25+ Monate	12,3	62,0	25,7	81
Insgesamt ( $R = 0$ )	58,3	26,8	14,9	258
Mikrozensuspanel insgesamt ( $D = 2$ )				
1 ≤12 Monate	87,4	4,0	8,6	585
2 13-24 Monate	70,6	18,1	11,3	504
3 25+ Monate	12,3	62,0	25,7	292
Insgesamt ( $R = 1, 0$ )	65,4	21,4	13,2	1.381
Beschäftigtenstichprobe und Mikrozensuspanel				
Insgesamt	64,2	22,3	13,5	24.451

Datenbasis: Tabelle 3.5. Modellspezifikation (Logit-Modell):  $Y|DX \{XY\} R|DXY \{RD RDX\}$ ;  
 Devianz  $G^2 = 50,04$ ; d.f. = 6; Log-Likelihood = -47.033,12; Anzahl log-linearer Parameter = 15.

Tabelle 3.7 zeigt, welche Verteilungen unter der Annahme nicht ignorierbarer Ausfälle im Mikrozensuspanel mit dem Modell  $Y|DX \{XY\} R|DXY \{RDY\}$  geschätzt werden.

Tabelle 3.7: Unter der Annahme nicht ignorierbarer Ausfälle (NINR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			
	1 Auszub.	2 Abschluss	3 Abbruch	Insgesamt
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Beschäftigtenstichprobe ( $D = 1$ )				
1 ≤ 12 Monate	87,3	4,0	8,7	9.950
2 13-24 Monate	70,5	18,1	11,5	7.706
3 25+ Monate	12,3	61,7	26,0	5.414
Insgesamt	64,1	22,2	13,7	23.070
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤ 12 Monate	89,8	4,1	6,2	486
2 13-24 Monate	73,2	18,5	8,3	414
3 25+ Monate	13,4	66,8	19,8	227
Insgesamt ( $R = 1$ )	68,3	22,0	9,7	1.127
ohne Angaben zum Ausbildungsstatus ( $R = 0$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤ 12 Monate	75,4	3,7	20,9	99
2 13-24 Monate	57,8	16,0	26,2	90
3 25+ Monate	8,1	44,0	47,9	65
Insgesamt ( $R = 0$ )	52,0	18,3	29,7	255
Mikrozensuspanel insgesamt ( $D = 2$ )				
1 ≤ 12 Monate	87,3	4,0	8,7	585
2 13-24 Monate	70,5	18,1	11,5	504
3 25+ Monate	12,3	61,7	26,0	292
Insgesamt ( $R = 1, 0$ )	65,3	21,3	13,4	1.381
Beschäftigtenstichprobe und Mikrozensuspanel				
Insgesamt	64,1	22,2	13,7	24.451

Datenbasis: Tabelle 3.5. Modellspezifikation (Logit-Modell):  $Y|DX \{XY\} R|DXY \{RD RDY\}$ ; Devianz  $G^2 = 51,03$ ; d.f. = 6; Log-Likelihood = -47.033,62; Anzahl log-linearer Parameter = 15; „boundary solution“.

Im Vergleich zum NINR-Modell, das ausschließlich auf Daten des Mikrozensuspanels (siehe Tab. 3.4) beruht, ergeben sich andere Verteilungen. Der wohl auffälligste Unterschied ist darin zu erkennen, dass nun bei den Ausfällen auch Anteile von Auszubildenden geschätzt werden und somit das vorherige unplausible Ergebnis nicht mehr auftritt.

Weil die Beschäftigtenstichprobe definitionsgemäß keine Ausfälle enthält, werden die NINR-Schätzungen (siehe Abschnitt 3.3.3) nur auf Basis der Odds der geschätzten Verteilungen der Personen ohne Ausfall des Mikrozensuspanels ( $D = 2, R = 1$ ) berechnet. Die Odds des Mikrozensuspanels und der Beschäftigtenstichprobe unterscheiden sich deshalb. Auch wenn für die Ausfallgruppe in Tabelle 3.7 keine Zellenbesetzungen mit Null erkennbar sind, liegt dennoch eine sogenannte „boundary solution“ vor, die damit zusammenhängt, dass mit der im Abschnitt

3.3.3 beschriebenen Lösung für Ausbildungsabsolventen negative Zellenbesetzungen geschätzt werden.

Die noch verbleibenden sechs Freiheitsgrade können dazu genutzt werden, mit dem Modell  $Y|DX \{XY\} R|DXY \{RD RDX RDY\}$  ein vermutlich angemesseneres Modell zu schätzen, das neben bedingt zufälligen, von der Beschäftigungsdauer abhängigen Ausfällen auch nicht ignorierbare, mit dem Ausbildungsstatus verbundene Ausfälle berücksichtigt. Die Devianz dieses Modells beträgt  $G^2 = 13,03$  (d.f. = 4; siehe Tab. 3.8). Sie ist deutlich niedriger als die Devianzen des MAR-Modells ( $G^2 = 50,04$ ) und des ersten NINR-Modells ( $G^2 = 51,03$ ). Da durch dieses Modell eine signifikant bessere Anpassung an die beobachteten Werte erreicht werden kann, ist es zu präferieren.

Betrachtet man die geschätzten Anteile des (nicht beobachteten) Ausbildungsstatus der Ausfallgruppe ( $R = 0$ ) in Tabelle 3.8, wird für 64 Prozent der Ausgefallenen der Status eines Auszubildenden geschätzt; dies entspricht etwa den Randverteilungen der beobachteten Daten und ist im Wesentlichen auf die Annahme bedingt zufälliger Ausfälle (MAR) zurückzuführen. Jedoch werden mit diesem Modell insgesamt nur knapp fünf Prozent der Ausfälle den Absolventen zugeordnet, die bei den beobachteten Daten der Beschäftigtenstichprobe rund 22 Prozent und im Mikrozensuspanel 25 Prozent umfassen.

Für 31 Prozent der Ausgefallenen wird geschätzt, dass sie die Ausbildung abgebrochen haben (siehe Tab. 3.8). Die entsprechenden Anteile liegen bei den beobachteten Daten ( $R = 1$ ) des Mikrozensuspanels bzw. bei der Beschäftigtenstichprobe mit neun bzw. 14 Prozent deutlich niedriger. Diese Schätzungen der Anteile der Absolventen und Ausbildungsabbrecher unter den Ausgefallenen konnten ansatzweise schon in dem NINR-Modell zum Mikrozensuspanel ohne Berücksichtigung der Beschäftigtenstichprobe (siehe Tab. 3.7) beobachtet werden. Sie sind somit im Wesentlichen auf die Modellierung nicht ignorierbarer Ausfälle zurückzuführen.

Das für die Ausfälle mit diesen einfachen Beispieldaten erzielte Modellergebnis lässt sich dahingehend interpretieren, dass erfolgreiche Ausbildungsabsolventen eine geringere, Ausbildungsabbrecher aber eine größere Ausfallwahrscheinlichkeit infolge räumlicher Mobilität aufweisen. Ob dieses vorläufige Ergebnis auch Bestand hat, wenn zusätzliche erklärende Variablen betrachtet werden, bleibt den späteren Analysen vorbehalten.

Zusammenfassend kann festgehalten werden, dass diese Modelle nicht nur die Prüfung substanzwissenschaftlicher Hypothesen zu Zusammenhängen in der Gesamttabelle (inkl. Ausfällen), sondern auch die Überprüfung von Annahmen über Zusammenhänge zwischen dem Ausfall und den Analysevariablen erlauben (Allison 2002; Baker und Laird 1988; Fay 1986; Little 1985; Molenberghs und Verbeke 2005; Schafer 1997; Toutenburg et al. 2004; Vermunt 1997a; Winship et al. 2002). Es handelt sich dabei um ein in der Praxis bewährtes Verfahren, das zudem mit dem Programm LEM leicht anzuwenden ist. Wie oben gezeigt, können dabei auch Verteilungen von Referenzdaten direkt in den Modellen berücksichtigt werden. In den Beispielen wurde nur eine abhängige Variable für den Übergang vom Ausbildungsstatus im Jahre 1996 auf den Ausbildungsstatus im Jahre 1997 ( $Y$ ) behandelt. Mit einer Erweiterung können auch weitere Übergänge (z. B.  $Y_t, Y_{t+1}$ ) und entsprechende zeitabhängige Ausfallzusammenhänge modelliert werden (Conaway 1992, 1993; Rendtel 1995: 274-281; siehe hierzu auch die Anwendung in Abschnitt 5.4.3).

Tabelle 3.8: Unter der Annahme bedingt zufälliger (MAR) und nicht ignorierbarer Ausfälle (NINR) geschätzter Ausbildungsstatus – Zeilenprozentwerte

Beschäftigungsdauer 1996 (X)	Ausbildungsstatus im April 1997 (Y)			
	1 Auszub.	2 Abschluss	3 Abbruch	Insgesamt
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Beschäftigtenstichprobe ( $D = 1$ )				
1 ≤12 Monate	87,4	4,0	8,6	9.950
2 13-24 Monate	70,7	17,9	11,4	7.706
3 25+ Monate	12,5	61,3	26,2	5.414
Insgesamt	64,2	22,1	13,7	23.070
mit Angaben zum Ausbildungsstatus ( $R = 1$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤12 Monate	87,6	4,6	7,7	495
2 13-24 Monate	68,9	21,3	9,7	418
3 25+ Monate	8,1	79,8	12,1	211
Insgesamt ( $R = 1$ )	65,8	24,9	9,3	1.124
ohne Angaben zum Ausbildungsstatus ( $R = 0$ )				
Mikrozensuspanel ( $D = 2$ )				
1 ≤12 Monate	86,4	0,3	13,4	90
2 13-24 Monate	79,1	1,4	19,5	86
3 25+ Monate	23,9	13,3	62,8	81
Insgesamt ( $R = 0$ )	64,2	4,8	31,0	257
Mikrozensuspanel insgesamt ( $D = 2$ )				
1 ≤12 Monate	87,4	4,0	8,6	585
2 13-24 Monate	70,7	17,9	11,4	504
3 25+ Monate	12,5	61,3	26,2	292
Insgesamt ( $R = 1, 0$ )	65,5	21,2	13,4	1.381
Beschäftigtenstichprobe und Mikrozensuspanel				
Insgesamt	64,3	22,0	13,7	24.451

Datenbasis: Tabelle 3.5. Modellspezifikation (Logit-Modell):  $Y|DX \{XY\} R|DXY \{RD RDX RDY\}$ ; Devianz  $G^2 = 13,03$ ; d.f. = 4; Log-Likelihood: -47.014,62; Anzahl log-linearer Parameter = 17; „boundary solution“.

## 4 Auszug aus dem Elternhaus

Im letzten Kapitel wurde das statistische Verfahren beschrieben, mit dem die Fragestellung potenziell selektiver Ausfälle infolge räumlicher Mobilität im Mikrozensuspanel bearbeitbar ist. Anhand der beispielhaft an bivariaten Tabellen beschriebenen Vorgehensweise ist deutlich geworden, wie wichtig a priori Kenntnisse über Ausfälle allgemein und insbesondere externe Referenzdaten für die Evaluation der Selektivitätsfrage sind. Während für die obigen beispielhaften Analysen einfache Ad-hoc-Vermutungen über mögliche Ausfallgründe als hinreichend gelten können, sollten aber die zu modellierenden Zusammenhänge bei den eigentlichen Ausfallanalysen explizit theoretisch begründet werden. Abgesehen von dem allgemeinen Postulat, dass die angemessene Erklärung eines Tatbestandes immer theoretisch fundiert sein sollte, sind Theorien auch von erheblichem praktischen Nutzen, da sie Hinweise auf die mit dem Ausfall verbundenen Merkmale geben und damit Annahmen über mögliche Ausfallgründe systematisch in die Modellierung einfließen können.

Vor diesem Hintergrund stellen Migrationstheorien einen allgemeinen theoretischen Rahmen für die mit räumlicher Mobilität verbundenen Ausfälle dar (vgl. Frick 1996; Kalter 1997, 2000; Wagner 1989: 173ff.). Bei der Prüfung, ob bei der Analyse von Bildungsverläufen Jugendlicher und junger Erwachsener selektive Stichprobenausfälle auftreten, ist zu beachten, dass die räumliche Mobilität in dieser Altersgruppe eng mit dem Auszug aus dem Elternhaus verbunden ist. Theoretische Ansätze zum Auszug aus dem Elternhaus bilden ein Teilgebiet von Migrationstheorien und liefern die für ein Mobilitätsmodell benötigten Informationen. Untersuchungen zum Auszugsverhalten können zwar nichts über den Status nach dem Auszug und potenzielle Verzerrungen sagen, die durch die Nichtberücksichtigung räumlich mobiler Personen entstehen können. Dennoch bieten diese Auswertungen über Stichprobenausfälle von Jugendlichen eine wichtige Aufklärung, die den nachfolgenden Analysen zur Stichprobenselektivität von Bildungsübergängen konzeptionell vorgelagert ist.

Das Kapitel ist wie folgt aufgebaut: Zunächst werden theoretische Konzepte zum Auszug aus dem Elternhaus dargestellt und der Forschungsstand skizziert. Darauf aufbauend werden im zweiten Abschnitt die Konstruktion von Variablen beschrieben sowie erste deskriptive Analysen zum Auszug aus dem Elternhaus bzw. Stichprobenausfall gezeigt. Der dritte Abschnitt enthält nach einer Beschreibung des für die Verlaufsdatenanalysen gewählten statistischen Modells die damit erzielten Ergebnisse. Abschließend werden auf Basis der Befunde vorläufige Schlussfolgerungen zur Stichprobenselektivität des Mikrozensuspanels gezogen.

### 4.1 Theoretische Konzepte und Forschungsstand

Für Jugendliche stellt der Auszug aus dem Elternhaus einen wichtigen Schritt des Übergangs in den von den Eltern unabhängigen Status der Erwachsenen dar. Während noch bis zu den 1960er-Jahren der Auszug eng mit der Eheschließung und Gründung einer eigenen Familie gekoppelt war, haben eine rückläufige Neigung zur Heirat, eng verbunden mit einer Lockerung der Normen über die Ehe als legitime Form des Zusammenlebens und der steigende Anteil von „neuen“ Formen des Zusammenlebens, wie beispielsweise nichteheliche Lebensgemeinschaften, zu einer Entkoppelung vom traditionellen Auszugsgrund einer Heirat geführt. Diese auch als Individualisierung und Diversifizierung von Lebensläufen bezeichnete Entwicklung muss jedoch nicht zugleich mit einer Destandardisierung von Auszugsmustern verbunden sein, da der Übergang ins

Erwachsenenalter in weitere altersspezifische Statuspassagen eingebettet ist (Konietzka und Huinink 2003). Dazu zählen die allgemein bildenden und beruflichen Ausbildungsetappen sowie der Einstieg ins Erwerbsleben.

Nach White (1994: 83-87) lassen sich Untersuchungen zum Auszugsverhalten hinsichtlich der hauptsächlich verwendeten theoretischen Konzepte nach drei Hauptrichtungen unterscheiden:

In der *Lebensverlaufsforschung* werden die Auszüge aus dem Elternhaus als ein Schritt altersspezifischer Entwicklungsetappen betrachtet, die weitgehend durch soziale Normen gesteuert werden.

Die *makrostrukturelle Perspektive* legt bei der Untersuchung der Entwicklung des Auszugsverhaltens den Schwerpunkt insbesondere auf Veränderungen ökonomischer und politischer Randbedingungen, wie beispielsweise die für Frauen durch den Zuwachs von Teilzeitarbeitsplätzen verbesserten Erwerbsgelegenheiten oder den Ausbau wohlfahrtsstaatlicher Leistungen (z. B. Ausbildungsförderung).

*Handlungstheoretische Ansätze* betrachten den Auszug aus dem Elternhaus als Ergebnis einer Entscheidung über die Lebensform einer Person – bei den Eltern vs. Allein- oder Zusammenleben mit anderen Personen –, die abhängig von den jeweils mit den Handlungsalternativen verbundenen Vor- und Nachteilen bzw. Nutzen und Kosten getroffen wird (einen Überblick über verschiedene Theorien rationalen Handelns geben Diekmann und Voss 2003).

Im Folgenden dient der handlungstheoretische Ansatz zur Orientierung bei der Darstellung auszugsrelevanter Zusammenhänge. Der Nutzen und die Kosten einer Handlungsalternative können sowohl monetäre als auch nicht-monetäre Aspekte umfassen, z. B. die Mietkosten einer eigenen Wohnung oder Freiheitseinschränkungen aufgrund elterlicher Kontrolle. Gleichwohl lassen sich die Beweggründe des Auszugs aus dem Elternhaus im Rahmen einer Sekundärdatenanalyse des Mikrozensus sicher nur näherungsweise und unvollständig als Nutzen und Kosten spezifizieren (vgl. zur Modellierung von Kosten, Nutzen und Präferenzen in der Sozialforschung u. a. Blossfeld und Prein 1998; Davidov et al. 2002; Opp 1999).

In der folgenden Darstellung werden in enger Anlehnung an Huinink und Konietzka (2000) und Konietzka und Huinink (2003) Merkmale der Jugendlichen und ihres Elternhauses beschrieben, die den Auszug beschleunigen oder hinausschieben. Zugleich wird dabei ein knapper Überblick zum Forschungsstand gegeben. Die Rahmenbedingungen der Entscheidung über den Auszug lassen sich grob nach den Dimensionen objektiver Gelegenheitsstruktur, eigenen Ressourcen der Jugendlichen und subjektiven Handlungsstrukturen unterscheiden.

Zur *objektiven Gelegenheitsstruktur* zählen gesellschaftliche sowie in der Familie liegende Bedingungen. Beispielsweise sind Jugendliche aus Wohnorten mit einer schwachen Wirtschaftsstruktur häufig gezwungen, aus Gründen der schulischen bzw. beruflichen Ausbildung oder zur Aufnahme einer Erwerbstätigkeit umzuziehen (Georg et al. 1994; Mulder et al. 2002).<sup>14</sup> Aufgrund der besonderen Arbeitsmarktprobleme in den neuen Bundesländern ist dort mit einer höheren Auszugsneigung zu rechnen. So zeigt die regionalisierte Sächsische Wanderungsanalyse, dass als Hauptgrund für Fortzüge aus Sachsen der Arbeitsplatz genannt wird (Statistisches Landesamt des Freistaates Sachsen 2003; vgl. auch insgesamt zur Ost-West-Migration Roloff 2001).

Ballungsräume bieten vielfältige Ausbildungs- und Arbeitsmöglichkeiten. Deshalb wäre für Personen aus kleineren Gemeinden tendenziell eine stärkere räumliche Mobilität als für Personen aus Ballungsräumen zu erwarten. Der häufig festgestellte umgekehrte Zusammenhang (Wagner 1989: 183) verweist jedoch auf zwischen Stadt und Land unterschiedliche Sozialisationsbedin-

14 In den Analysen von Mulder et al. (2002) ist der Effekt jedoch nur für eine Gemeindegröße von 25-50 Tsd. Einwohnern und bei Frauen positiv.



gungen und Lebensentwürfe von Jugendlichen. Zu beachten sind aber auch die mit der Gemeindetypisierung korrelierten räumlich verschiedenen Verhältnisse von Mietwohnungen und Hausbesitz (Georg et al. 1994).

In Bezug auf die sozioökonomische Lage der Familie ist einerseits zu erwarten, dass finanziell besser gestellte Familien die Jugendlichen bei der Gründung eines eigenen Haushalts leichter unterstützen können. So ermittelte Ermisch (1996) für Studenten positiv mit dem Elterneinkommen verbundene Auszugswahrscheinlichkeiten. Andererseits können Familien mit höherem Einkommen ihren Kindern häufig ein komfortableres Wohnumfeld bieten als einkommensschwächere Familien (Iacovou 2001; Lauterbach und Lüscher 1999), sodass diese Jugendlichen länger zu Hause wohnen. Mit Ausnahme von Jugendlichen, die aus Ausbildungsgründen (teilweise temporär) von zu Hause ausziehen, dürften insgesamt betrachtet höhere Familien- bzw. Haushaltseinkommen die Auszugsneigung verringern.

In Retrospektivstudien wird mangels Angaben zum Haushaltseinkommen auf sozioökonomische Variablen der Eltern als Proxy-Variablen zurückgegriffen. Entsprechende Analysen von Mulder et al. (2002), Weick (1993) sowie Ziegler und Schladt (1993) ergeben nur schwache Zusammenhänge zwischen dem sozioökonomischen Status der Familie und dem Auszugsverhalten. Werden jedoch Einkommensangaben direkt verwendet, ändert sich der Befund. Lauterbach und Lüscher (1999) finden starke positive Zusammenhänge zwischen dem Haushaltseinkommen und dem Auszugsalter. Des Weiteren zeigen ihre Analysen, dass Kinder von Landwirten (insbes. die ältesten Söhne) spät ausziehen bzw. als potenzielle Nachfolger des Betriebes im Elternhaus wohnen bleiben. Ähnlich ermittelt Gartner (2000) für Jugendliche, die nach dem Auszug in einen Mehrpersonenhaushalt ziehen, eine mit steigendem Haushaltseinkommen sinkende Auszugswahrscheinlichkeit. Bei den nach Geschlecht getrennten Analysen zeigen sich signifikante Effekte aber nur für Frauen. Für Personen, die nach dem Auszug in Einpersonenhaushalten leben, sind die Einkommenseffekte statistisch nicht signifikant.

Das Auszugsalter variiert auch mit der Familiengröße, und zwar in der Weise, dass Jugendliche mit mehreren Geschwistern das Elternhaus früher verlassen. Dieser Zusammenhang lässt sich auf die bei größeren Familien beengteren Wohnverhältnisse zurückführen (Wagner 1989; Lauterbach und Lüscher 1999).

Hinsichtlich des Bildungsniveaus des Elternhauses zeigen sich empirisch positive Zusammenhänge mit der Auszugswahrscheinlichkeit, d. h. Jugendliche aus formal höher gebildeten Elternhäusern haben eine höhere Auszugsneigung (Georg et al. 1994; Mulder et al. 2002; Weick 1993). Aufgrund der Korrelationen zwischen dem Bildungsniveau und der Einkommenslage des Elternhauses sowie mit der Bildungsbeteiligung der Jugendlichen sind Effekte des Bildungsniveaus der Eltern aber schwierig von anderen Faktoren zu trennen. Zudem dürfte das elterliche Bildungsniveau eng mit Normen oder Werthaltungen verbunden sein: „It may be that parents who have completed higher education themselves expose their children to different norms and expectations whilst they are growing up, not only encouraging further education but also the delay of other more traditional markers, such as marriage, in favour of education“ (Hillman und Marks 2002: 5).

Migranten in Deutschland weisen im Vergleich zu Deutschen stärker traditionelle Familienkulturen auf (Nauck 2000; Weidacher 2000). Jugendliche Migranten heiraten zwar früher, ziehen aber meistens erst aus diesem Anlass aus dem Elternhaus aus (vgl. White (1994) für ähnliche Befunde aus den USA zu afroamerikanischen und lateinamerikanischen Jugendlichen). Ein Grund sind vermutlich die stärkeren ökonomisch-utilitaristischen Erwartungen an intergenerative Beziehungen, z. B. dass Kinder ihren Eltern im Alter helfen sollten, die Nauck (2000: 364) insbesondere bei türkischen und vietnamesischen Familien festgestellt hat. Für diese Gruppen ist daher ein späterer Auszug aus dem Elternhaus als bei Deutschen zu erwarten. Betrachtet man Auslän-



der insgesamt und alle Umzüge ohne Berücksichtigung von Haushaltszusammenhängen, zeigt aber eine Wanderungsanalyse für Baden-Württemberg (Stein 2004), dass Ausländer vor allem aufgrund von Umzügen in das Herkunftsland deutlich räumlich mobiler sind als Deutsche. Werden nur Umzüge innerhalb des Bundeslandes berücksichtigt, unterscheiden sich männliche jugendliche Ausländer nur geringfügig von deutschen Jugendlichen, während ausländische Frauen im jungen Erwachsenenalter seltener als deutsche Frauen umziehen.

Weitere Handlungsgelegenheiten und Handlungsrestriktionen sind durch die mit den altersspezifischen Statuspassagen verbundenen Lebensbedingungen gegeben. Besonders die Übergänge zwischen dem Bildungs- und Erwerbssystem stellen zentrale Ereignisse im Lebenslauf dar, die eng mit der Verfügbarkeit über eigene Ressourcen (siehe unten) verbunden sind. Im Zusammenhang mit der Bildungsexpansion ab den 1960er-Jahren wird vermutet, dass der im Vergleich zu früher längere Schulbesuch zu einem längeren Verbleib im Elternhaus beigetragen hat (Eggen 1999: 193) und damit dem längerfristigen Trend der Verringerung des Auszugsalters entgegenläuft (vgl. auch Lauterbach und Lüscher 1999).

Jugendliche können aus Gründen der schulischen oder beruflichen Ausbildung vollständig oder vorübergehend von zu Hause ausziehen; teilweise werden eigene Wohnungen (z. B. am Studienort) parallel zum Wohnen bei den Eltern unterhalten. Nach dem Abschluss der Ausbildungsphase und mit der Aufnahme einer Erwerbstätigkeit ist verstärkt ein Auszug aus dem Elternhaus zu erwarten. Ergebnisse von Georg et al. (1994), deren Untersuchung diese lebenslaufbezogenen Variablen berücksichtigt hat, zeigen aber weder für den Abschluss einer Berufsausbildung noch für den Berufseintritt statistisch signifikante Effekte auf die Auszugswahrscheinlichkeit. Dagegen finden Konietzka und Huinink (2003: 302) deutliche Belege für einen im Zeitverlauf enger gewordenen Zusammenhang zwischen Ausbildungsbeginn und Berufseinstieg mit dem Auszug aus dem Elternhaus. Es ist aber auch davon auszugehen, dass Jugendliche nicht sofort nach dem Übergang von der Ausbildung ins Erwerbsleben von zu Hause ausziehen, sondern erst nachdem eine sichere Beschäftigung und eine stabile Einkommenssituation erreicht sind (Huinink und Konietzka 2000: 8).

Für die jüngeren Geburtskohorten (1971-78) berichten Konietzka und Huinink (2003: 300) eine Differenz zwischen dem Medianalter beim Ausbildungsbeginn und beim Auszug von etwa vier Jahren für Männer und zwei bis rund zweieinhalb Jahren für Frauen. Diese geschlechtsspezifischen Unterschiede können auf unterschiedliche Altersnormen der Partnerwahl und Familiengründung zurückzuführen sein (Georg et al. 1994). Zu beachten sind aber auch Wehr- und Zivildienstzeiten der Männer, die für einen längeren Verbleib im Elternhaus verantwortlich sein können.

Neben diesen Zusammenhängen sind die eigenen Ressourcen der Jugendlichen von Bedeutung. Insbesondere ist zu erwarten, dass mit einem höheren eigenen Erwerbseinkommen die Auszugswahrscheinlichkeit steigt. So zeigt die Untersuchung von Georg et al. (1994), dass ein für den Lebensunterhalt als ausreichend wahrgenommenes eigenes Einkommen mit einer höheren Auszugsneigung einhergeht.

Nicht zuletzt spielen, wie bereits zum Bildungsniveau des Elternhauses erwähnt, psychologische und soziale Dispositionen als *subjektive Handlungsstrukturen* der jungen Erwachsenen eine Rolle. Dazu zählen die emotionalen Beziehungen zu Eltern, Werthaltungen gegenüber der Familie oder Kirche bzw. Religion und die Vorstellungen über die eigene Lebensplanung. Es wird angenommen, dass Kinder, die nur mit einem Elternteil aufwachsen, z. B. sogenannte Scheidungswaisen, weniger enge Beziehungen zu den Eltern entwickeln, in diesem Familientyp weniger „soziales Kapital“ zur Verfügung steht und die Kinder deshalb früher ausziehen als Jugendliche,

die mit beiden leiblichen Eltern zusammenleben (Juang et al. 1999: 506; White 1994: 91).<sup>15</sup> Dagegen finden Lauterbach und Lüscher (1999) für Kinder verwitweter Elternteile, tendenziell auch für geschiedene Elternteile, einen längeren Verbleib im Elternhaus, den sie im Zusammenhang mit Bindungen zu den Eltern bzw. Unterstützungen der Eltern durch die Kinder sehen.

Wie oben bereits angesprochen, haben sich die Gründe für das Verlassen des Elternhauses im Zeitverlauf erheblich vom früher wichtigsten Grund „Heirat“ bzw. „Gründung einer eigenen Familie“ in Richtung „Unabhängigkeit“ verändert. Dementsprechend finden Konietzka und Hui-nink (2003: 303) im Zeitverlauf, d. h. für die Geburtskohorten 1929-31 bis 1959-61, einen deutlichen Rückgang des Doppelereignisses „Ehe“ und „Auszug“.

Zusammenfassend lassen sich folgende Arbeitshypothesen formulieren:

Die Auszugsneigung ist stärker

- in den neuen Bundesländern,
- in Großstädten als in kleinen Gemeinden,
- bei Frauen,
- bei eigenem (Erwerbs-) Einkommen der Jugendlichen,
- nach Abschluss der allgemein bildenden bzw. beruflichen Ausbildungsphase,
- bei finanziell schlechter gestellten Familien,
- bei größeren Familien bzw. bei Jugendlichen mit mehreren Geschwistern,
- bei formal höher gebildeten Eltern und
- bei allein erziehenden Elternteilen.

## 4.2 Abgrenzung der Stichprobe und Variablenkonstruktion

Welche der oben skizzierten theoretisch begründeten Zusammenhänge und Kovariaten sich mit den Mikrozensusdaten abbilden lassen, wird im Unterabschnitt 4.2.2 dargestellt. Zunächst sind Fragen der Abgrenzung der Stichprobe zu klären, weil die Ergebnisse auch davon abhängen, wie die Analyseeinheiten ausgewählt werden und wie das Ereignis „Auszug“ definiert werden (White 1994).

### 4.2.1 Abgrenzung der Stichprobe

Die Darstellung des Auszugsverhaltens von Jugendlichen und jungen Erwachsenen auf Basis des Mikrozensuspanels konzentriert sich auf die Phase ab dem Ende der Pflichtschulzeit bis zu den zum Zeitpunkt der ersten Panelbefragung (1996) 26-Jährigen, da in diesem Alter in der Regel die allgemeine und berufliche Ausbildung beendet ist. Bei der Abgrenzung der Altersgruppe ist auch zu berücksichtigen, dass für multivariate Analysen ausreichende Fallzahlen verfügbar sind.

Für Auszugsanalysen stehen im Mikrozensus verschiedene Angaben zur Verfügung. Auf Haushaltsebene liegen für jeden Befragten Angaben zum Verwandtschaftszusammenhang zur Haushaltsbezugsperson vor (siehe Heidenreich und Nöthen 2002; Lengerer et al. 2005). Auf Basis der Fragen zu nichtehelichen Lebensgemeinschaften (ohne Auskunftspflicht) lässt sich ermitteln, ob Jugendliche und junge Erwachsene Kinder der Haushaltsbezugsperson oder dessen Lebenspartners sind. Beim Verwandtschaftszusammenhang zur Haushaltsbezugsperson wird allerdings nicht zwischen Schwieger- und leiblichen Söhnen bzw. Töchtern sowie Schwiegervater

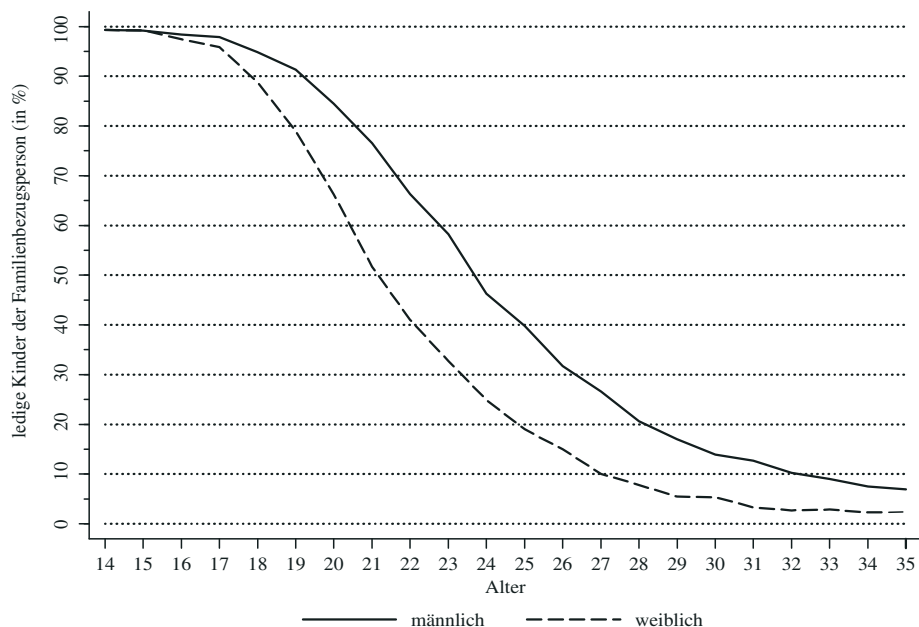
---

15 Eventuell ziehen Kinder geschiedener Eltern aber auch nur von einem zum anderen Elternteil.

bzw. -mutter differenziert. Bei Mehrfamilienhaushalten kann damit nicht entschieden werden, wer Kind der Haushaltsbezugsperson ist, bzw. wer noch im elterlichen Haushalt lebt oder Schwiegertochter bzw. -sohn ist. Aufgrund dieser Probleme wird für die Abgrenzung der im Elternhaus lebenden Kinder im Folgenden die Angabe zur Stellung in der Familie des sogenannten traditionellen Familienkonzepts verwendet, mit der ledige Kinder in der Familie erfasst werden.

Im Längsschnitt ist bei Mehrfamilienhaushalten zu beachten, dass die Familiendefinition von der zeitlich konsistenten und korrekten Zuordnung der im Haushalt befragten Personen zu Familien abhängt. Diese Zuordnung wird nicht erfragt, sondern in den statistischen Landesämtern mithilfe der Angaben zum Verwandtschaftsverhältnis zur Haushaltsbezugsperson vorgenommen. Dies ist allerdings nicht eindeutig möglich, wenn Haushalte keine oder mehrere Ehepaare enthalten (Stauder 2002). Die bereits bei Querschnittauswertungen beobachteten teilweise unplausiblen Angaben zum Verwandtschaftsverhältnis bei jungen Personen (Stauder 2002: 25) treten auch bei Auswertungen des Mikrozensuspanels auf.

Ergänzend zur Querschnittsanalyse zeigen die Paneldaten, dass eine Reihe von Familienzuordnungen fehlerhaft werden, wenn die Haushaltsbezugsperson im Zeitverlauf wechselt (siehe Schimpl-Neimanns 2006d: 43). Dieses Problem tritt insbesondere dann auf, wenn Befragte ihre Verwandtschaftsbeziehung als „sonstige Verwandte“ angeben. Nach Dateninspektionen wird der Umfang dieser potenziell unplausiblen Fälle im Mikrozensuspanel zwar nur auf zwei Prozent geschätzt, es erscheint aber ratsam, bei Längsschnittauswertungen die Veränderungen der Beziehungen zur Familienbezugsperson vorläufig nicht zu analysieren, zumal zurzeit noch keine längsschnittlich konsistenten Familiennummern verfügbar sind. Dies heißt aber auch, dass die Auswahl der Analysetichprobe bzw. Risikopopulation des Jahres 1996 in geringem Umfang auf fehlerbehafteten Angaben zur Stellung zur Familienbezugsperson beruhen kann.



Quelle: Mikrozensus 1996 (faktisch anonymisierte 70 %-Substichprobe); Bevölkerung in Privathaushalten am Hauptwohnsitz; hochgerechnete und an die Bevölkerungsfortschreibung angepasste Fallzahlen (Frauen n = 12.427.000; Männer n = 11.738.000).

Abbildung 4.1: Anteil der Personen, die 1996 bei ihren Eltern leben

Abbildung 4.1 zeigt auf Basis der Querschnittsdaten des Mikrozensus 1996, dass am Ende der Pflichtschulzeit praktisch noch alle Jugendlichen bei ihren Eltern leben und Frauen früher als Männer von zu Hause ausziehen. Von den 18-jährigen Frauen wohnen noch 89 Prozent, von den 18-jährigen Männern 95 Prozent zu Hause. Das mittlere Auszugsalter (Median) von Frauen liegt bei 21 Jahren, bei Männern beträgt es 24 Jahre. Während von den 35-jährigen Männern noch sieben Prozent als lediges Kind in ihrer Familie leben, sind es bei den Frauen nur zwei Prozent.

In die Analysestichprobe des Mikrozensuspanels wurden Personen übernommen, die 1996 15 bis 26 Jahre alt waren und als lediges Kind in ihrer Familie lebten (siehe Tab. 4.1). Für diese Gruppe wird untersucht, bis zu welchem Zeitpunkt sie aus der Erhebung infolge eines Wegzugs aus dem ausgewählten Haushalt ausscheidet.

Aus der Risikopopulation wurden Personen ausgeschlossen, für die keine plausiblen Zusammenführungsergebnisse vorliegen. Des Weiteren wurden jene Jugendliche nicht berücksichtigt, die bei der Befragung von 1996 bis 1999 wenigstens einmal ausgefallen sind, dann aber wieder befragt wurden (temporäre Ausfälle).<sup>16</sup> Damit werden zwar Aspekte des mehrmaligen Auszugs bzw. der Rückkehr ins Elternhaus ausgeblendet, jedoch sind die Fallzahlen für eine eigenständige Analyse dieses Mobilitätstyps zu gering.<sup>17</sup>

Tabelle 4.1: Fallauswahl beim Mikrozensuspanel für die Analysen zum Auszug aus dem Elternhaus

Population	Kinder <sup>*)</sup>	Sonst.	Insges.
1996 befragte 15- bis 26-jährige Personen, die zur Bevölkerung in Privathaushalten am Hauptwohnsitz zählen	10.361	6.003	16.364
- temporäre Ausfälle 1996-1999	90	67	157
- nicht plausibel zusammenführbare Personensätze	499	357	856
= Auswahl 1	9.772	5.579	15.351
- Haushaltszusammenführung auf Personenebene inkonsistent	512		
= Auswahl 2	9.260		
- Wegzug des gesamten Haushalts 1996-1999	949		
= Auswahl 3: Risikopopulation für Verlaufsanalysen	8.311		

\*) Ledige Kinder der Familienbezugsperson, Bevölkerung am Familienwohnsitz.

Quelle: Mikrozensuspanel 1996-1999 (faktisch anonymisierte 65 %-Substichprobe); nicht hochgerechnete Fallzahlen; eigene Berechnungen.

Weil die Auszugsanalysen auf Informationen zum Haushalts- bzw. Familienkontext aufbauen, kommen hierfür nur Jugendliche aus vollständig plausibel zusammengeführten Haushalten in Frage. Die resultierenden Selektionen reduzieren die Ausgangsstichprobe auf 9.260 Personen (Auswahl 2).

- 16 Vor diesem Hintergrund des Ausschlusses unplausibler Fälle sowie temporärer Ausfälle und in Anbetracht der Zusammenführungsprobleme bei der Erstellung des Panelfiles stellt sich die kritische Frage, wie gut das Auszugsverhalten mit der so abgegrenzten Risikopopulation (Auswahl 2 in Tab. 4.1) abgebildet werden kann. Erfreulicherweise ergaben aber Überprüfungen für den Zeitpunkt 1996 nahezu identische geschlechtsspezifische Verteilungen des Querschnitts- und Panelfiles mit einer Gesamtabweichung von weniger als einem Prozentpunkt.
- 17 Vgl. in diesem Kontext auch die Kritik von Wagner und Huinink (1991) an früheren Auszugsanalysen des Mikrozensus durch Schwarz (1989). Gartner (2000: 12) berichtet auf Basis des Sozioökonomischen Panels für diese Wechsel einen Anteil von 0,8 Prozent der Erstauszüge.

Schließlich ist zu beachten, dass für Jugendliche, die zwischen 1996 und 1999 mit dem gesamten Haushalt wegziehen,<sup>18</sup> nicht festgestellt werden kann, ob sie danach weiter mit ihren Eltern zusammenleben oder nicht. Reduziert man deshalb die Stichprobe weiter um die betreffenden Fälle, stehen für die Analyse 8.311 Personen zur Verfügung (Auswahl 3 in Tab. 4.1).

In Tabelle 4.2 sind die Veränderungen der Stichprobe insgesamt sowie der Risikopopulation lediger Kinder von Familienbezugspersonen infolge von Weg- und Zuzügen zwischen 1996 und 1997 dargestellt.

*Tabelle 4.2:* Veränderungen der Risikopopulation der 15- bis 26-jährigen Personen im Jahre 1996 durch Weg- und Zuzüge zwischen 1996 und 1997

	1996 Insges. <sup>1)</sup>	darunter Kinder <sup>2)</sup> (%)	1997 Insges. <sup>1)</sup>	darunter Kinder <sup>2)</sup> (%)
Gesamthaushalt immobil	11.355	70,5	11.381	70,0
Wegzug von Personen aus weiter befragten Haushalten	977	83,8		
Wegzug mit ganzem Haushalt	1.674	26,3		
<i>Wegzüge zusammen</i>	<i>2.651</i>	<i>47,5</i>		
Zuzug von Personen in bereits 1996 befragten Haushalten			351	49,3
Zuzug ganzer Haushalte			2.000	13,5
<i>Zuzüge zusammen</i>			<i>2.351</i>	<i>18,8</i>
Insgesamt (n)	14.006	9.260	13.732	8.406
(in %)	(100,0)	(66,1)	(100,0)	(61,2)

<sup>1)</sup> Bevölkerung in Privathaushalten am Hauptwohnsitz im Alter von 15 bis 26 Jahren im Jahre 1996.

<sup>2)</sup> Ledige Kinder der Familienbezugsperson, Bevölkerung am Familienwohnsitz (Zeilenprozentwerte).

*Quelle:* Mikrozensuspanel 1996-1999 (siehe Tab. 4.1); eigene Berechnungen.

Der Vergleich des Anteils lediger Kinder (Risikopopulation) der 1996 mit dem ganzen Haushalt fortgezogenen Jugendlichen (26,3 %) mit dem Anteil lediger Kinder der 1997 mit dem ganzen Haushalt neu Hinzugezogenen (13,5 %) deutet darauf hin, dass schätzungsweise rund die Hälfte der betreffenden Jugendlichen nach dem Umzug nicht mehr bei den Eltern lebt. Eine altersspezifische Analyse (hier ohne Darstellung) zeigt, dass dies überwiegend für Jugendliche ab 19 Jahren zutrifft.

#### 4.2.2 Variablenkonstruktion

Im Folgenden werden die Merkmale des Mikrozensuspanels in Anlehnung an die Übersicht der theoretisch relevanten Rahmenbedingungen der individuellen Entscheidungen von Huinink und Konietzka (2000: 4) dargestellt. Die meisten Variablen beziehen sich auf das Startjahr 1996 des Mikrozensuspanels. Für sie werden pragmatischerweise zunächst zeitkonstante Effekte auf das Auszugsverhalten angenommen. Um aber die mit dem Auszug verbundenen Faktoren zu ermitteln, sind zusätzlich Merkmale zu berücksichtigen, die im Zeitverlauf den Auszug fördern oder

18 Eine theoretisch angemessene Modellierung dieses Typs räumlich mobiler Jugendlicher wäre auf Haushaltsebene durchzuführen. Zumindest teilweise wären dafür andere Erklärungsfaktoren notwendig (vgl. Frick 1996; Kalter 2000).

hemmen. Hierfür werden für verschiedene Bereiche hypothesengeleitet zeitveränderliche Variablen konstruiert, mit denen Effekte der bis zum Auszug beobachteten Statusveränderungen auf das zeitabhängige Ereignis des Auszuges geprüft werden können. Tabelle A.1 im Anhang enthält deskriptive Verteilungen zu diesen Variablen.

#### *Objektive Gelegenheitsstruktur:*

##### *Lokale Infrastruktur des Wohnortes und der lokalen Lebensbedingungen*

Für die Erfassung von Aspekten der lokalen Infrastruktur stehen im Mikrozensuspanel das Bundesland, die Gemeindegrößenklasse und der siedlungsstrukturelle Gemeindetyp zur Verfügung. Da sich die Auszugsmuster zwischen dem früheren Bundesgebiet und den neuen Bundesländern unterscheiden (vgl. Huinink und Konietzka 2000; Juang et al. 1999), sind getrennte Analysen sinnvoll.

Mit der Variablen Gemeindegrößenklasse sind Stadt-Land-Unterschiede nur eingeschränkt zu untersuchen, da beispielsweise die Lebensverhältnisse kleiner Gemeinden im ländlichen Raum sehr verschieden von denen im verstädterten Raum oder in unmittelbarer Nähe einer Großstadt sein können. Die Bevölkerungsdichte, der Grad der Urbanisierung und die Zentralität bzw. die räumliche Verdichtung von Gemeinden sind aber mit dem im Arbeitsfile verfügbaren siedlungsstrukturellen Gemeindetyp (Böltkens 1997; Böltkens und Irmen 1997) typisiert worden. Explorative Analysen zeigen, dass die 18 Kategorien der Ursprungsvariablen ohne größeren Informationsverlust in drei Klassen zusammengefasst werden können. Der Gemeindetyp (Variable *R*) umfasst somit Kernstädte (kreisfreie Städte mit über 100.000 Einwohnern), Ober- / Mittelzentren (Städte und Gemeinden mit zentralörtlichen Funktionen) und sonstige Gemeinden.

##### *Eigenschaften der Familie: Einkommen und Wohnsituation*

Zur Charakterisierung der Einkommenssituation der Jugendlichen kann das Haushaltseinkommen herangezogen werden. Mit der Berechnung des bedarfsgewichteten verfügbaren Haushaltsnettoeinkommens (Äquivalenzeinkommen) pro Haushaltsmitglied wird die Haushaltsstruktur, d. h. die Haushaltsgröße und das Alter der Personen, berücksichtigt (vgl. Stauder und Hüning 2004).<sup>19</sup> Um eventuelle nichtlineare Zusammenhänge mit dem Auszugsverhalten zu kontrollieren, wurde zunächst das Äquivalenzeinkommen in etwa zehn gleich stark besetzte Gruppen kategorisiert und fehlende Einkommensangaben in einer zusätzlichen Kategorie ausgewiesen.<sup>20</sup> Tests zeigen, dass die Kategorien der Einkommensvariablen (*E*) in fünf Klassen zusammengefasst werden können: unter 948 DM, 948-1.359 DM, 1.359-1.890 DM, 1.890-2.903 DM, 2.903 und mehr DM, sowie ohne Einkommensangabe.

Die ökonomische Situation der Eltern ist eng verbunden mit dem Wohnungs- und Hausbesitz. Die Analyse der Mikrozensus-Zusatzerhebung zeigt, dass von den 1998 noch befragten Jugendlichen, deren Eltern Wohnungseigentümer sind, mit circa 13 Prozent deutlich weniger ausfallen bzw. wegziehen als in der Gruppe der zur Miete wohnenden Personen (23 %). Da die Angaben zur Wohnsituation nur in der 1998 durchgeführten Zusatzerhebung vorliegen und die Wohnungsidentifikatoren des Arbeitsfiles im Längsschnitt nicht konsistent sind, muss aber auf die Differenzierungen verzichtet werden.

19 Die Gewichte zur Ermittlung des bedarfsgewichteten Pro-Kopf-Haushaltsnettoeinkommens orientieren sich an der alten OECD-Skala, die für die erste Person im Haushalt ein Gewicht von Eins, für weitere Personen 0,7 und für Kinder unter 15 Jahren 0,5 vorsieht.

20 Etwa jeweils die Hälfte der fehlenden Angaben von insgesamt rund sechs Prozent hängt mit Antwortverweigerungen zusammen, beziehungsweise betrifft Haushalte, in denen wenigstens ein Haushaltsmitglied selbstständiger Landwirt ist, da in dieser Gruppe keine Einkommensangaben erfasst werden.



### *Familienstruktur*

Von den im Mikrozensus vorhandenen Faktoren zur Beschreibung der Strukturen des Zusammenlebens wird der Familientyp (Variable  $F$ ) herangezogen. Neben Ehepaaren mit ledigen Kindern werden Alleinerziehende mit ledigen Kindern (Verwitwete, Geschiedene und verheiratete, nicht zusammenlebende Elternteile) klassifiziert. Um zu kleine Fallzahlen zu vermeiden, wurden alle Kategorien von allein erziehenden Elternteilen zusammengefasst.

Daneben wird die Anzahl der Geschwister (Variable  $K$ ) in der Familie zum Zeitpunkt 1996 ermittelt. Hierbei werden alle ledigen Kinder der Familienbezugsperson ohne Altersbegrenzung berücksichtigt. Drei und mehr Geschwister werden in einer Kategorie zusammengefasst.

### *Erwerbsbeteiligung und sozialer Status der Eltern*

Die Erwerbsbeteiligung und der soziale Status der Eltern werden anhand der Angaben der Familienbezugsperson erfasst. Bei vollständigen Familien ist dies in der Regel der Vater. Unter den im Mikrozensus hierfür relevanten Variablen wurde die sozioökonomische Variable „Stellung im Betrieb“ gewählt. Verschiedene Testauswertungen zeigen, dass die differenzierten Angaben zusammengefasst werden können, sodass letztlich nur die sozialrechtlichen Kategorien der beruflichen Stellung der Familienbezugsperson (Variable  $S^{FB}$ ) zu berücksichtigen sind: Nichterwerbstätige, Selbstständige (einschl. mithelfende Familienangehörige), Beamte, sowie Arbeiter und Angestellte (zusammengefasst).

Der allgemeine und berufliche Schulabschluss der Familienbezugsperson (Variable  $C^{FB}$ ) wird mithilfe der CASMIN-Bildungsskala (Brauns und Steinmann 1999; Granato 2000) umgesetzt. Aus Fallzahlgründen wurden verschiedene Zusammenfassungen vorgenommen. Die Variable enthält folgende Kategorien: Hauptschule ohne beruflichen Abschluss und kein Abschluss, Hauptschule mit beruflichem Abschluss, Mittlere Reife, Fachhochschulreife oder Abitur, Fachhochschulabschluss, Hochschulabschluss, keine Angaben zum Bildungsabschluss der Familienbezugsperson.

### *Individuelle Ressourcen und Lebenslage der Jugendlichen: Ausbildung und Erwerbsbeteiligung*

Einige bisherige Untersuchungen zeigen, dass der Auszug aus dem Elternhaus in starkem Maße mit dem Ausbildungs- und Berufsverlauf der Jugendlichen zusammenhängt (siehe S. 46). Der Besuch allgemein bildender sowie beruflicher Schulen im Jahre 1996 (Variable  $B$ ) wird mit folgenden Ausprägungen erfasst: allgemein bildende Schule, berufliche Schule, Fachhochschule, Hochschule, kein Schulbesuch.

Da davon auszugehen ist, dass die Auszugsneigung nach Abschluss der Ausbildungsphase steigt, wird das 1996 bereits erreichte Bildungsniveau der Jugendlichen (Variable  $C$ ) als Kontrollvariable mithilfe der CASMIN-Bildungsskala konstruiert (Brauns und Steinmann 1999; Granato 2000). Wie Testauswertungen zeigen, können die Kategorien jedoch ohne Informationsverlust in Bezug auf das Auszugsverhalten wie folgt zusammengefasst werden: noch Schul-/Hochschulbesuch, höchstens Hauptschule ohne berufliche Ausbildung, Hauptschule mit beruflicher Ausbildung oder Mittlere Reife oder Abitur, Fachhochschule oder Hochschule, keine Bildungsangabe vorhanden.

Die Erwerbsbeteiligung der Jugendlichen wird mittels der Angaben zur betrieblichen Stellung abgebildet. Ähnlich wie bei der Einstufung der sozioökonomischen Lage der Familienbezugspersonen können verschiedene Zusammenfassungen vorgenommen werden, die auch zur Vermeidung kleiner Fallzahlen notwendig sind. Die berufliche Stellung der Jugendlichen (Variable  $S$ )

wird wie folgt abgegrenzt: Nichterwerbstätige, Selbstständige, Beamte (einschl. Soldaten), Arbeiter, Angestellte, Auszubildende, Wehr- und Zivildienstleistende.

Wie oben bereits erwähnt, ist zu erwarten, dass neben den Angaben zum Ausgangszeitpunkt 1996 die Übergänge vom Bildungs- ins Erwerbssystem wichtige Veränderungen der auszugsrelevanten Faktoren erfassen. Aus diesem Grund wird ermittelt, ob zwischen 1996 und 1997 sowie 1996 bis 1997 vs. 1998 der Schulbesuch endet (Variablengruppe V). Daneben werden Wechsel vom Besuch einer allgemein bildenden Schule zum Besuch einer beruflichen Schule (Variablengruppe W) berücksichtigt, da diese Übergänge mit räumlicher Mobilität (z. B. zur Aufnahme einer Berufsausbildung im dualen System) und teilweise eigenem Einkommen verbunden sein können. Auch diese Informationen beschreiben die Statusübergänge zwischen 1996 und 1997 sowie zwischen 1996 bis 1997 vs. 1998. Unter der Annahme, dass Arbeitsplatzmobilität und räumliche Mobilität zusammenhängen, werden außerdem die Retrospektivangaben über Betriebswechsel im Jahr vor der jeweils aktuellen Erhebung verwendet (Variablengruppe Z; Betriebswechsel 1995/96, 1996/97, 1997/98). Um eventuelle zeitverzögerte Effekte zu erfassen, wird, wenn ein Wechsel stattfand, der neue Wert der zeitveränderlichen Indikatorvariablen auch bei späteren Zeitpunkten beibehalten.

### *Eigenes Einkommen*

Zur Erfassung des Niveaus der ökonomischen Unabhängigkeit von den Eltern wird die Information über die Quelle des überwiegenden Lebensunterhalts (Variable L) mit folgenden Zusammenfassungen verwendet: (1) Eltern, Rente, Vermögen oder Pflegeversicherung, (2) Erwerbstätigkeit oder Arbeitslosengeld/-hilfe, (3) Sozialhilfe oder sonstige Unterstützung (z. B. Bafög). In inhaltlicher Hinsicht werden teilweise heterogene Einkommensquellen zusammengefasst. Diese Vergrößerungen sind notwendig, um zu kleine Zellenbesetzungen zu vermeiden. Dabei ist aber eine möglichst hohe Ähnlichkeit der Gruppen in Bezug auf das Auszugsverhalten gewährleistet. Z. B. unterscheiden sich Jugendliche mit Arbeitslosengeld oder -hilfe als überwiegender Quelle des Lebensunterhalts nicht wesentlich von Jugendlichen, die ihren Unterhalt aus eigener Erwerbstätigkeit bestreiten.

Zeitliche Veränderungen der Quelle des überwiegenden Lebensunterhalts von den Eltern zur (eigenen) Erwerbstätigkeit zwischen 1996 und 1997 sowie zwischen 1997 und 1998 werden mit der Variablengruppe U abgebildet. Auch in diesen Variablen wird ein Wertwechsel bei späteren Übergängen beibehalten (s. o.).

Grundsätzlich wäre es wichtig, zusätzlich die Höhe des persönlichen Nettoeinkommens der Jugendlichen zu verwenden. Dies ist jedoch einerseits schwierig, weil diese Einkommensteile bereits im Haushaltseinkommen bzw. im bedarfsgewichteten Äquivalenzeinkommen (Variable E) enthalten sind. Andererseits korreliert das persönliche Nettoeinkommen auch mit den Variablen Lebensunterhaltsquelle (L), Schulbesuch (B), Bildungsniveau (C) und berufliche Stellung (S). Die Verwendung des persönlichen Nettoeinkommens würde daher zu Multikollinearitätsproblemen führen. Da sich in Tests eine statistisch größere Erklärungskraft in Bezug auf das Auszugsverhalten für das Äquivalenzeinkommen und die anderen Variablen zeigte, wird auf das persönliche Nettoeinkommen verzichtet.

### *Demografische Angaben*

Schließlich wird das Geschlecht (Variable G) verwendet, das sich in bisherigen Untersuchungen als stärkster Prädiktor des Auszugs erwiesen hat. Neben dem Lebensalter, das in den Verlaufsdatenanalysen die Prozesszeit bildet, wird das Alter zum Zeitpunkt 1996 (Variable A) berücksichtigt, um Kohorteneffekte zu kontrollieren.



Im Unterschied zu vielen anderen Studien, die sich auf Deutsche beschränken, kann mit dem Mikrozensus auch nach der Staatsangehörigkeit differenziert werden. Um bei den multivariaten Analysen zu kleine Fallzahlen zu vermeiden, wird die Staatsangehörigkeit (Variable  $N$ ) nicht nach einzelnen Migrantengruppen, sondern lediglich nach Deutschen und EU-Ausländern einerseits und sonstigen Ausländern andererseits untergliedert. Die Gruppe der sonstigen Ausländer setzt sich zu rund 58 Prozent aus türkischen Jugendlichen und zu 19 Prozent aus Jugendlichen aus dem ehemaligen Jugoslawien zusammen.

Jugendliche mit einem zweiten Wohnsitz haben bereits den partiellen Auszug aus dem Elternhaus vollzogen. Da zu erwarten ist, dass diese Personen, die am Ort der (elterlichen) Hauptwohnung erfasst sind und einen Nebenwohnsitz haben, ein größeres Ausfall- bzw. Auszugsrisiko besitzen, wird das Merkmal Haupt- / Nebenwohnsitz (Variable  $H$ ) verwendet.<sup>21</sup>

### 4.3 Verlaufsdatenanalysen

Zur Ermittlung der partiellen Effekte der einzelnen erklärenden Variablen werden im Folgenden Verlaufsdatenanalysen durchgeführt. Zunächst ist zu diskutieren, welches statistische Modell für die Daten des Mikrozensuspanels geeignet ist.

#### 4.3.1 Das diskrete Hazardraten-Regressionsmodell

In der Regel stellt in statistischen Modellen zum Auszug aus dem Elternhaus das Alter der Personen die Zeitachse bzw. Prozesszeit dar. Für die Verweildauer im Elternhaus wird eine diskrete Zufallsvariable  $T$  definiert, deren Werte  $j$  den Zustandswechsel von  $y_{ij} = 0$  „bei den Eltern wohnend“ zu  $y_{ij} = 1$  „Auszug aus dem Elternhaus“ indizieren. Während der Auszug zu jedem Zeitpunkt erfolgen kann, liegen im Mikrozensus durch das Erhebungsdesign bedingt nur Beobachtungen zwischen April 1996 und April 1999 im Abstand von einem Jahr vor. Das statistische Verfahren sollte deshalb berücksichtigen, dass die Zustandswechsel nur innerhalb diskreter Zeitintervalle  $j = [a_{t-1}, a_t)$  beobachtet werden können. Des Weiteren ist zu beachten, dass über die Zeitpunkte vor 1996 keine Informationen vorhanden sind. Das Fehlen von Angaben zum Zeitpunkt vor April 1996 wird als Linksstützung („left-truncation“; Guo 1993) bezeichnet. Außerdem ist nicht bekannt, ob und wann Personen nach 1999 ausziehen. Es handelt sich somit sowohl um linksgestützte (vor 1996) als auch rechtszensierte (nach 1999) Daten.

Die Hazardrate  $h(t_{ij})$  entspricht im diskreten Fall der bedingten Wahrscheinlichkeit, dass eine Person  $i$  im Altersabschnitt  $j$  ( $T_i = j$ ) aus dem Elternhaus auszieht, gegeben, dass bis dahin noch kein Auszug stattgefunden hat:

$$h(t_{ij}) = P(T_i = j \mid T_i \geq j).$$

Der Zustandswechsel kann als logistische Regression geschätzt werden, wenn aus den Personendaten, in denen jede Person mit einem Satz enthalten ist, ein sogenannter Personen-Perioden-Datensatz erstellt wird. Dieser enthält eine Beobachtung für jedes Zeitintervall  $j$ , in dem eine Person bis zum Ereignis des Auszuges bzw. bis zur Zensierung 1999 im Panel verblieben ist, und daneben den Indikator des Ereignisses  $y_{ij}$  im Zeitintervall.<sup>22</sup>

21 Während in der Analysestichprobe insgesamt sieben Prozent der Jugendlichen einen Zweitwohnsitz haben, beträgt dieser Anteil bei Studenten rund 31 Prozent.

22 Die dabei entstehende Verletzung der Annahme unabhängig identisch verteilter Beobachtungen kann bei der Modellierung korrigiert werden (siehe Rogers 1993; Wooldridge 2002: 405ff.).

Die Koeffizienten des Modells setzen sich aus Regressionskonstanten  $\alpha_j$  für die einzelnen Zeitintervalle und Steigungskoeffizienten  $\beta$  zusammen. Bezeichnet man den Kovariatenvektor mit  $\mathbf{x}_{ij}$ , ist die mit der logistischen Regression geschätzte diskrete Hazardrate:

$$h(t_{ij}) = 1 / (1 + \exp(-\alpha_j - \beta \mathbf{x}_{ij}))$$

Für Verlaufsanalysen wird oft das Regressionsmodell von Cox verwendet (siehe u. a. Blossfeld et al. 1986). Im Cox-Modell unterscheiden sich die Hazardraten verschiedener Werte der erklärenden Variablen nur durch eine multiplikative Konstante von der Basisrate; sie sind also proportional zueinander. Für die mittels logistischer Regression geschätzten Hazardraten trifft diese Eigenschaft jedoch nicht zu.<sup>23</sup> Im Unterschied zum Cox-Modell stehen hier nicht die Hazardraten, sondern deren Kreuzprodukt (Odds-Ratio; kurz OR) für jeden Zeitpunkt in einem proportionalen Verhältnis:

$$OR = \frac{h(t_{ij} | x = 1) / (1 - h(t_{ij} | x = 1))}{h(t_{ij} | x = 0) / (1 - h(t_{ij} | x = 0))} = \exp(\beta' \mathbf{x}_{ij})$$

Die Überlebensfunktion im Ausgangszustand zu verbleiben, d. h. die Wahrscheinlichkeit am Ende des Altersabschnitts  $j$  noch bei den Eltern zu wohnen, unter der Bedingung, dass bis dahin noch kein Auszug stattgefunden hat, ist:

$$S(t_j) = S(t_{j-1}) [1 - h(t_j)]$$

Analog zur üblichen logistischen Regression ist die Log-Likelihood des diskreten Hazardraten-Regressionsmodells:

$$l = \sum_{j=1}^n \sum_{i=1}^J y_{ij} \ln(h(t_{ij})) + (1 - y_{ij}) \ln(1 - h(t_{ij}))$$

Im Folgenden wird als Maß der Modellanpassung die Devianz verwendet (siehe Abschnitt 3.3.1, Seite 31). Die in der Likelihood-Ratio-Teststatistik berechnete Differenz der Devianzen verschiedener Modelle ist Chi-Quadrat-verteilt. Zur Beurteilung der Modellanpassung dient zusätzlich Aikakes Informationskriterium  $AIC = G^2 + 2 * (\text{Anzahl der Parameter})$ , das auch bei nicht-geschachtelten Modellen verwendet werden kann. Als „bestes“ Modell ist das Modell mit dem niedrigsten AIC-Wert zu präferieren (Singer und Willett 2003: 121-122).

Bei dieser Modellierung des Auszugsverhaltens beziehen sich die zeitunabhängigen Angaben zumeist auf das 15. Lebensjahr der Befragten. Mit zeitabhängigen Variablen wird beispielsweise erfasst, ob eine Person vor dem Auszug in der Ausbildung oder erwerbstätig war (siehe z. B. Mayer und Wagner 1986). Als Prozesszeit dient i. d. R., wie bereits erwähnt, das Lebensalter. Hinsichtlich des Zeitbezugs der erklärenden Variablen und in Bezug auf die Prozesszeit weicht das Mikrozensuspanel von den sonst bei Auszugsanalysen verwendeten Daten ab, da nur für das Zeitfenster 1996 bis (maximal) 1999 Beobachtungen vorhanden sind. In Tabelle 4.3 ist der Ver-

23 Im Rahmen des allgemeinen linearen Modells kann alternativ zu dem hier verwendeten Modell (Logit-Verbindungsfunktion; proportional odds model) mittels komplementärer log-logistischer Verbindungsfunktion ein Modell geschätzt werden, bei dem die Hazardraten proportional zueinander sind (proportional hazards model). Aufgrund der diskreten Zeitangaben im Mikrozensus wird aber die einfachere logistische Regression verwendet (siehe Singer und Willett 2003: 425).

bleib im Mikrozensuspanel nach dem Lebensalter und der Kalenderzeit als zweidimensionales Lexis-Diagramm dargestellt.

*Tabelle 4.3:* Verbleib 15- bis 26-jähriger Jugendlicher des Jahres 1996 im Elternhaus nach Lebensalter und Erhebungszeitpunkt

Lebensalter	Erhebungszeitpunkt			
	1996 (n)	1997 (in %)	1998 (in %)	1999 (in %)
(29)				(46,2)
(28)			(61,7)	(47,0)
27		77,6	60,0	47,8
26	303	78,2	60,5	52,6
25	413	79,3	64,5	56,7
24	458	79,6	69,4	60,8
23	504	85,6	74,7	62,6
22	520	86,5	75,6	67,3
21	613	88,4	78,8	71,8
20	692	89,0	84,4	80,6
19	825	92,6	89,8	88,3
18	949	96,1	94,2	94,2
17	964	98,4	97,6	
16	1.012	99,1		
15	1.058			
Insgesamt (n)	8.311			
in Prozent	100	90,1	80,4	70,1

*Quelle:* Mikrozensuspanel 1996-1999; eigene Berechnungen (siehe Tab. 4.1).

So wohnen von den beim Start des Panels 15-Jährigen ( $n = 1.058$ ) drei Jahre später noch 94 Prozent im Elternhaus. Dies trifft bei den 26-Jährigen ( $n = 303$ ) nur auf 46 Prozent der Befragten zu. Die unterschiedliche Altersverteilung ergibt sich daraus, dass die Risikopopulation nur die 1996 noch im Elternhaus wohnenden ledigen Kinder der Familienbezugsperson umfasst. Unterscheidet man nicht danach, ob eine Person noch zuhause wohnt, sind die Alterskohorten näherungsweise gleich groß.

Mit wenigen Ausnahmen werden im Mikrozensus keine Retrospektivfragen gestellt, sodass über die Zeitpunkte vor 1996 keine Informationen vorliegen. Wegen dieser Linksstützung ist es nicht möglich, Variablen zu konstruieren, die sich für alle Altersgruppen auf den gleichen Zeitpunkt im Lebensverlauf beziehen (z. B. Status der Eltern, als die Person 15 Jahre alt war). Die erklärenden Variablen enthalten dagegen Informationen zum Zeitpunkt 1996 bzw. bei zeitveränderlichen erklärenden Variablen Informationen zum Zeitpunkt vor einem Statuswechsel.

Um den direkten Bezug der Angaben auf den Erhebungszeitpunkt zu erhalten, kommt als alternative Modellierung die Kalenderzeit als Prozesszeit sowie das Alter im Jahre 1996 als Kontrollvariable für Kohorteneffekte in Frage. Da jedoch die Hazardraten erheblich stärker vom Lebensalter statt vom Erhebungszeitpunkt abhängen, ist davon abzuraten (siehe Korn et al. 1997; Singer und Willett 2003).

Stattdessen sollten Verlaufsdatenanalysen mit dem Lebensalter als Prozesszeit durchgeführt werden. Hierfür ist es nur nötig, das Alter beim Eintritt in die Risikopopulation zu berücksichtigen, sodass die Prozesszeit mit dem Alter im Jahre 1996 startet (Singer und Willett 2003: 319, 595). Jedoch könnte die heterogene altersspezifische Zusammensetzung der Risikopopulation damit verbunden sein, dass sich die im Startjahr 1996 noch zu Hause wohnenden älteren Jugendlichen als Spätauszieger („Nesthocker“) von den anderen Gruppen unterscheiden (siehe Lauterbach und Lüscher 1999). Um solche eventuellen Effekte des unterschiedlichen Eintrittsalters zu kontrollieren, wird die Alterskohorte 1996 (Variable A) als erklärende Variable in das Modell aufgenommen.

Um kleine Zellenbesetzungen bei den älteren Jugendlichen zu vermeiden, werden die Personen in den folgenden Verlaufsdatenanalysen bis zum Alter von 27 Lebensjahren betrachtet. Personen, die bis zu diesem Zeitpunkt noch nicht ausgezogen sind, werden als zensierte Fälle behandelt. In Tabelle 4.3 sind die so ausgeklammerten Übergänge gekennzeichnet. Wie man anhand der ähnlichen Verbleibwahrscheinlichkeiten benachbarter Zellen sehen kann, sind durch dieses Vorgehen keine Verzerrungen zu erwarten.

### 4.3.2 Modellergebnisse

Tabelle 4.4 zeigt für das frühere Bundesgebiet (Westdeutschland inkl. West-Berlin) und die neuen Bundesländer (Ostdeutschland) getrennte Analysen des Auszugsverhaltens. Neben dem Modell, mit dem lediglich Konstanten für die diskreten Lebensaltersintervalle ohne erklärende Variablen geschätzt werden (Modell 1), sind die Ergebnisse eines Grundmodells (M2) dargestellt, in dem die Effekte aller zeitkonstanten und zeitveränderlichen Variablen enthalten sind. In den Modellen 3 bis 20 wird jeweils eine Variable entfernt (Rückwärtsselektion). Die Likelihood-Ratio-Statistik zeigt summarisch den partiellen statistischen Einfluss dieser Variablen auf das Auszugsrisiko.<sup>24</sup>

Die zur Kontrolle von potenziellen Kohorteneffekten verwendete Alterskohorte 1996 (Variable A; M4) ist statistisch nicht signifikant ( $P = 0,82$ ). Somit kann davon ausgegangen werden, dass keine Kohorteneffekte vorliegen und im Modell alle Alterseffekte durch das Lebensalter als Prozesszeit absorbiert sind.

Sowohl in West- als auch in Ostdeutschland sind die AIC-Werte bei Herausnahme des Geschlechts (Variable G; M5) am größten, d. h. die Zusammenhänge zwischen Geschlecht und Auszugswahrscheinlichkeit sind am stärksten. Der zweitstärkste partielle Effekt ist ebenfalls in beiden Gebieten für den Gemeindetyp (Variable R; M3) festzustellen.

Bei allen weiteren Erklärungsfaktoren des Auszugsverhaltens unterscheiden sich jedoch westdeutsche von ostdeutschen Jugendlichen deutlich. Konzentriert man sich auf die stärksten und statistisch signifikanten Zusammenhänge, ist in Westdeutschland, neben Geschlecht und Gemeindetyp, vor allem der Schulbesuch (Variable B; M8) bedeutsam. Danach folgen mit etwa gleicher Erklärungsleistung (AIC) der Abschluss von Schulbesuchsetappen (V; M17), Familientyp (F; M10), Betriebswechsel (Z; M20), berufliche Stellung der Familienbezugsperson ( $S^{FB}$ ; M12), Schularartwechsel (W; M18), erreichtes Bildungsniveau (C; M16), Geschwisterzahl (K; M11) und

24 Aus Platzgründen enthält Tabelle 4.4 keine Freiheitsgrade (d.f.) für Ostdeutschland; siehe hierzu die entsprechenden Spalten für Westdeutschland.

das Äquivalenzeinkommen ( $E$ ; M9). Dagegen sind für ostdeutsche Jugendliche, außer dem Geschlecht und dem Gemeindetyp, im Vergleich zum Grundmodell 2 nur die Geschwisterzahl ( $K$ ; M11) und ihre 1996 erreichte berufliche Stellung ( $S$ ; M14) statistisch signifikant.

Bemerkenswerterweise sind in beiden Teilen Deutschlands sowohl die Quelle des überwiegenden Lebensunterhalts im Jahre 1996 (Variable  $L$ ; M14) als auch deren Änderungen im Zeitverlauf (Variablengruppe  $U$ ; M19) im Vergleich zum Grundmodell 2 statistisch unbedeutend. Aber bedeutet das auch, dass aus einem Wechsel von der finanziellen Abhängigkeit von den Eltern zur überwiegenden Bestreitung des Lebensunterhalts aus eigenem Erwerbseinkommen keine Veränderungen der Auszugsneigung resultieren?

Teilweise wurde in früheren Untersuchungen ein Anstieg der Auszugsrate bei Beginn der Erwerbstätigkeit festgestellt; die Zusammenhänge sind aber im Zeitverlauf, d. h. bei jüngeren Kohorten, schwächer geworden (Mayer und Wagner 1986: 31). Konietzka und Huinink (2003: 308) fassen neuere Ergebnisse ebenfalls in dieser Richtung als Lockerung des Zusammenhangs zwischen dem Auszug und der Erlangung ökonomischer Unabhängigkeit von den Eltern zusammen. In diesen Kontext können auch die hier mit dem Mikrozensuspanel ermittelten Ergebnisse eingeordnet werden. Jedoch sind die Gründe für die schwachen partiellen Zusammenhänge der Lebensunterhaltsquelle vor allem darin zu suchen, dass diese Variablen (Variable  $L$  und Variablengruppe  $U$ ) eng mit dem Schulbesuch ( $B$ ), Änderungen der Bildungsbeteiligung ( $V$ ) und der beruflichen Stellung ( $S$ ) zusammenhängen. Statusübergänge von der allgemein bildenden und beruflichen Ausbildung zur Aufnahme einer regulären Erwerbstätigkeit gehen aber mit der Verfügbarkeit über eigene finanzielle Ressourcen einher. Entfernt man die mit dem Lebensunterhalt korrelierten Variablen aus dem Modell, zeigen sich sehr wohl signifikante Zusammenhänge zwischen dem Auszugsverhalten und der überwiegenden Quelle des Lebensunterhalts.

Dass die Variable Staatsangehörigkeit ( $N$ ; M6) keinen signifikanten Erklärungsbeitrag leistet, weist darauf hin, dass sich Migranten aus Nicht-EU-Ländern in Bezug auf das Auszugsverhalten nicht von Deutschen und EU-Ausländern unterscheiden. Dies ist aufgrund anderer Teilbefunde nicht ganz unerwartet (siehe S. 45).

Entgegen den Erwartungen (siehe S. 54) unterscheiden sich Jugendliche, die bereits eine eigene Wohnung am Nebenwohnsitz (Variable  $H$ ; M7) haben, nicht von anderen Jugendlichen. Gleichmaßen und im Unterschied zu anderen Untersuchungen (siehe S. 45) zeigen sich beim Bildungsniveau der Familienbezugsperson ( $C^{FB}$ ; M12) keine statistisch signifikanten Zusammenhänge mit dem Auszug aus dem Elternhaus.

Nimmt man die nicht signifikanten Variablen aus dem Grundmodell 2 heraus, ergeben sich für West- und Ostdeutschland sehr unterschiedliche Modelle. Um den Vergleich zu erleichtern, verbleiben deshalb im Modell 21 auch jene Variablen, die nur in einem Landesteil statistisch signifikant sind. Jedoch wird für Westdeutschland das bereits erreichte Bildungsniveau der Jugendlichen (Variable  $C$ ) aus dem Modell genommen, da trotz eines insgesamt signifikanten Erklärungsbeitrages kein einzelner Koeffizient dieser Variablen statistisch signifikant ist. Dies dürfte damit zusammenhängen, dass mit dem Schulbesuch ( $B$ ) sowie den zeitveränderlichen Bildungsvariablen (Variablengruppen  $V$  und  $W$ ) bereits wesentliche Zusammenhänge im Modell enthalten sind. Tabelle 4.4 zeigt, dass sich das resultierende Modell 21 nicht signifikant vom Grundmodell 2 unterscheidet. Die im Folgenden dargestellten Ergebnisse beziehen sich deshalb auf dieses Modell 21. Die Regressionskoeffizienten hierzu sind in Tabelle A.2 zu finden.

Tabelle 4.4: Ergebnisse logistischer Regressionen zum Auszug aus dem Elternhaus

Westdeutschland (inkl. West-Berlin)													Ostdeutschland			
M	Modellterme	Devianz		Likelihood-Ratio-Statistik			Devianz		Likelihood-Ratio-Statistik							
		G <sup>2</sup>	df	AIC	ΔG <sup>2</sup>	df	P	G <sup>2</sup>	AIC	ΔG <sup>2</sup>	P					
1	Zeitkonstanten α <sub>j</sub> [Lebensalter – 15; j = 1, ..., 12]	10.493,52	12	10.518				3.199,25	3.223							
2	+ R, A, G, N, H, B, E, F, S <sup>FB</sup> , C <sup>FB</sup> , L, S, C, V, W, U, Z	10.143,65	70	10.284	349,87	58	0,00	2.960,86	3.101	238,40	0,00					
Rückwärtsselektion: M2 - zeitkonst. Variablen (Status 1996)																
3	– R Gemeindetyp	10.178,52	68	10.315	34,87	2	0,00	2.992,85	3.129	31,99	0,00					
4	– A Alterskohorte	10.150,39	59	10.268	6,74	11	0,82	2.973,99	3.092	13,14	0,28					
5	– G Geschlecht	10.274,87	69	10.413	131,23	1	0,00	3.049,48	3.187	88,62	0,00					
6	– N Staatsangehörigkeit	10.144,04	69	10.282	0,39	1	0,53	2.962,72	3.101	1,86	0,17					
7	– H Haupt-/Nebenwohnsitz	10.147,25	69	10.285	3,60	1	0,06	2.960,87	3.099	0,02	0,90					
8	– B Schulbesuch	10.161,11	66	10.293	17,46	4	0,00	2.969,36	3.101	8,50	0,07					
9	– E Äquivalenzeinkommen	10.156,67	65	10.287	13,02	5	0,02	2.970,98	3.101	10,12	0,07					
10	– F Familientyp	10.150,84	69	10.289	7,19	1	0,01	2.964,26	3.102	3,40	0,07					
11	– K Geschwisterzahl	10.153,40	67	10.287	9,76	3	0,02	2.981,44	3.115	20,59	0,00					
12	– S <sup>FB</sup> Berufl. Stellung Familienbezugsperson	10.154,69	67	10.289	11,04	3	0,01	2.962,39	3.096	1,54	0,67					
12	– C <sup>FB</sup> Bildung Familienbezugsperson	10.150,70	64	10.279	7,05	6	0,32	2.972,37	3.100	11,51	0,07					
14	– L Lebensunterhaltsquelle	10.145,96	68	10.282	2,31	2	0,32	2.966,09	3.102	5,23	0,07					
15	– S Berufliche Stellung	10.154,45	65	10.284	10,80	5	0,06	2.977,28	3.107	16,43	0,01					
16	– C Bildungsniveau	10.155,42	66	10.287	11,77	4	0,02	2.967,63	3.100	6,77	0,15					
Rückwärtsselektion: M2 - zeitveränderliche Variablen																
17	– V Beendigung Schulbesuch 96/97, 96-97/98	10.153,48	68	10.289	9,83	2	0,01	2.964,66	3.101	3,81	0,15					
18	– W Schularwechsel 96/97, 96-97/98	10.152,10	68	10.288	8,45	2	0,02	2.961,27	3.097	0,41	0,81					
19	– U Wechsel Lebensunterh. 96/97, 96-97/98	10.143,93	68	10.280	0,28	2	0,87	2.961,04	3.097	0,19	0,91					
20	– Z Betriebswechsel 95/96, 96/97, 97/98	10.154,76	67	10.289	11,11	3	0,01	2.967,27	3.101	6,42	0,09					
21	– (A, N, H, E, F, C <sup>FB</sup> , L, C, U)	10.176,55	43	10.263	32,90	27	0,20	2.996,70	3.083	35,84	0,12					

In Bezug auf die Frage, ob die Auszugswahrscheinlichkeiten von Jugendlichen in Ostdeutschland höher als in Westdeutschland sind bzw. ob sich die Modellschätzungen beider Landesteile unterscheiden, zeigen Wald-Tests der Koeffizienten des gewählten Modells 21 auf Gleichheit zwischen West- und Ostdeutschland, dass dies insgesamt der Fall ist ( $\chi^2 = 119,75$ ; d.f. = 43;  $P = 0,00$ ). Jedoch trifft dies bei den einzelnen Variablen, einschließlich der Zeitkonstanten ( $\alpha_j$ ) mit Ausnahme der Merkmale Geschlecht und Geschwisterzahl nicht zu. Die Unterschiede beim Wald-Test betragen nach Geschlecht:  $\chi^2 = 8,92$  bei d.f. = 1, und nach der Geschwisterzahl:  $\chi^2 = 13,72$  bei d.f. = 3. Schätzt man das Modell 21 für beide Landesteile gemeinsam, ergibt sich für Ostdeutschland mit einem Odds-Ratio von 1,36 ein signifikant höheres Auszugsrisiko als in Westdeutschland. Berücksichtigt man die beim Wald-Test gefundenen Ergebnisse und schätzt zusätzlich Interaktionen für das Geschlecht und die Geschwisterzahl, ist der Haupteffekt für Ostdeutschland nicht mehr statistisch signifikant. Die Ost-West-Unterschiede sind somit vorwiegend auf die mit diesen beiden Variablen verbundenen Sachverhalte zurückzuführen. Für die Beschreibung und Diskussion der Einzelergebnisse wird das Modell 21 und die Unterscheidung nach den Landesteilen beibehalten.

Da es sich bei allen Variablen um qualitative, dummycodierte Variablen handelt, werden insgesamt 43 Koeffizienten geschätzt. Es erscheint daher sinnvoll, die Ergebnisse grafisch aufzubereiten und hierbei die Hazardraten zu betrachten. Analog zu den Abschnitten 4.1 und 4.2 werden die Befunde themenspezifisch diskutiert, jedoch orientiert sich die Reihenfolge an der Stärke der partiellen Effekte.

### 4.3.3 Diskussion der Ergebnisse

In den folgenden Grafiken sind jeweils die Hazardraten der Referenzkategorie der einzelnen erklärenden Variablen gleich. Hierbei ist zu beachten, dass für die Referenzgruppe wegen den entsprechenden Definitionen stark unterdurchschnittliche Auszugswahrscheinlichkeiten geschätzt werden.<sup>25</sup> Diese Auszugsrisiken kumulieren sich aber je nach Merkmalskombination, sodass bei der Darstellung einzelner Koeffizienten infolge der Definition der Referenzkategorie leicht der Eindruck geringer Auszugswahrscheinlichkeiten entstehen kann. Dies wird in Abbildung 4.2 deutlich, in der die empirischen Hazardraten des Modells ohne Kovariaten (Modell 1) den durch das Modell 21 für die Referenzkategorie geschätzten Hazardraten gegenübergestellt werden.

Die empirische Hazardrate in Abbildung 4.2 zeigt für die hier analysierte Stichprobe auch, dass die Auszugswahrscheinlichkeit in Ostdeutschland bei Jugendlichen im Alter von 19 Jahren stark ansteigt. In Westdeutschland ist bei 23- bis 24-Jährigen ein steilerer Anstieg zu erkennen. Mit dem Mikrozensuspanel kann zwar der Zielzustand nach dem Wegzug, der als ein Indikator von Auszugsgründen betrachtet werden kann, nicht ermittelt werden, doch ist in dieser Hinsicht auffällig, dass der Anstieg der Hazardrate in Ostdeutschland in etwa mit dem Alter beim Abschluss allgemeiner und beruflicher Ausbildungen zusammenfällt. In Westdeutschland korrespondiert der Anstieg mit dem mittleren Heiratsalter (Median) von 23 Jahren in dieser ausgewählten Altersgruppe der Risikopopulation.

25 Eine andere Abgrenzung der Referenzkategorien würde sich nur in einer Skalenverschiebung niederschlagen, die relativen Abstände zwischen den Gruppen würden gleich bleiben. Die Referenzgruppe ist wie folgt definiert: Gemeindetyp = sonstige Gemeinde; Geschlecht = männlich; Schulbesuch = allgemein bildende Schule; Pro-Kopf-Äquivalenzznettoeinkommen = unter 948 DM; Familientyp = Ehepaar mit ledigen Kindern; Geschwisterzahl = keine Geschwister; Berufliche Stellung der Familienbezugs-person = Arbeiter und Angestellte; Berufliche Stellung = nichterwerbstätig; alle zeitveränderlichen Variablen: „keine Veränderung“.



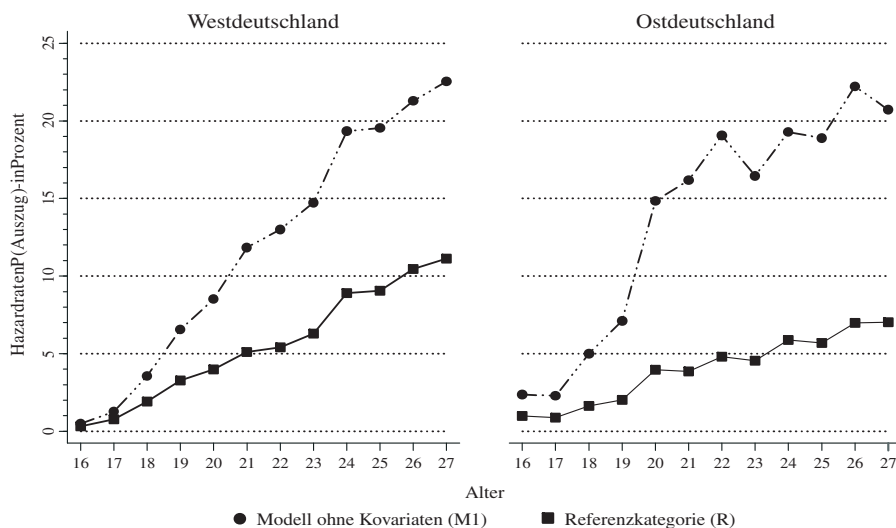


Abbildung 4.2: Hazardraten zum Auszug aus dem Elternhaus zu einem Modell ohne Kovariaten und für die Referenzkategorie (Modell 21)

Des Weiteren ist zu beachten, dass sich die durch das Modell geschätzten Hazardraten nicht direkt auf die Population beziehen lassen. So liefert das statistische Modell unter anderem auch Schätzungen bzw. Hazardraten zum Auszug für 26-jährige Besucher allgemein bildender Schulen oder für 16-jährige Personen mit Besuch einer Hochschule. Dennoch kann diese Darstellung hilfreich sein, insofern beispielsweise der Vergleich der Hazardrate für Besucher einer allgemein bildenden Schule mit der Hazardrate für Hochschulüler darüber informiert, welche veränderten Auszugswahrscheinlichkeiten bei einem Wechsel von einer allgemein bildenden Schule zur Hochschule durch das Modell prognostiziert werden.

#### Demografische Angaben: Geschlecht

Abbildung 4.3 ist zu entnehmen, dass Frauen eine erheblich höhere Auszugsneigung aufweisen als Männer. Diese spiegeln sich im Odds-Ratio der Hazardraten bei der hier verwendeten Abgrenzung der Referenzkategorie wider, d. h. den Wahrscheinlichkeitsverhältnissen von Frauen im Vergleich zu Männern, zu einem Zeitpunkt auszuziehen vs. nicht auszuziehen. Das Odds-Ratio der Hazardraten von Frauen vs. Männern ist in Westdeutschland fast doppelt so hoch ( $\exp(b) = 1,9$ ), in Ostdeutschland liegt es beim 2,7-fachen (siehe Spalte „Effektkoeffizient“ in Tabelle A.2). Diese Geschlechtsunterschiede sind in den neuen Bundesländern signifikant stärker als in Westdeutschland (siehe Seite 60).



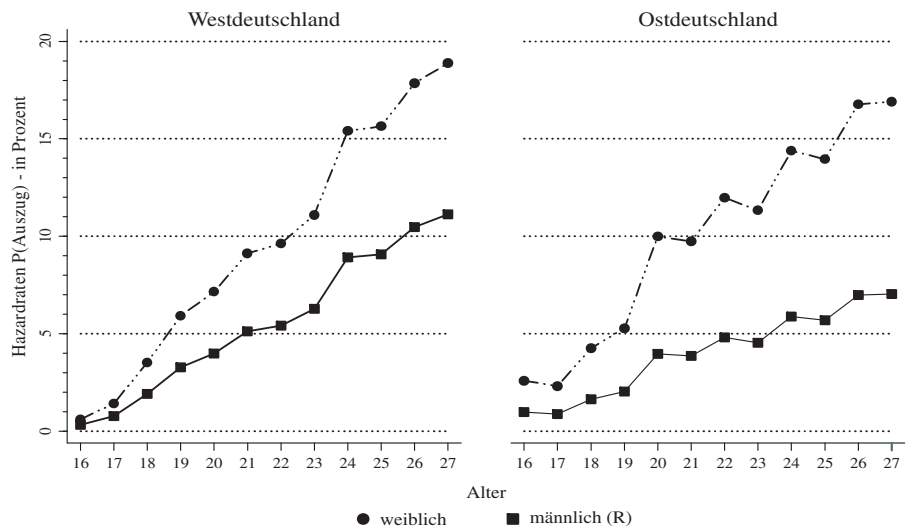


Abbildung 4.3: Hazardraten zum Auszug aus dem Elternhaus nach Geschlecht

Objektive Gelegenheitsstruktur: Lokale Infrastruktur des Wohnortes

Betrachtet man zu diesem Themenbereich die Modellergebnisse für den Indikator Gemeindetyp (siehe Abb. 4.4), zeigen sich wie erwartet für Jugendliche aus Kernstädten sowie Ober-/Mittelzentren deutlich höhere Auszugsneigungen als für Jugendliche im ländlichen Raum (Sonstige Gemeinden).

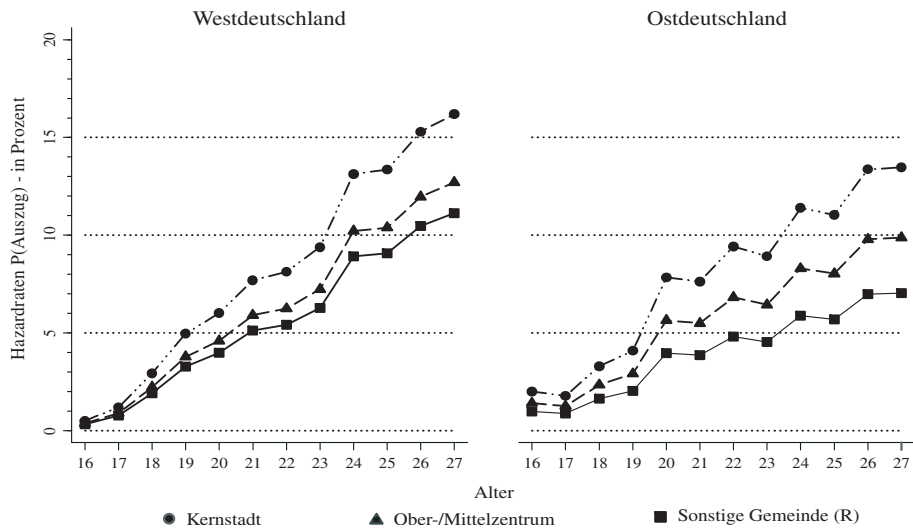


Abbildung 4.4: Hazardraten zum Auszug aus dem Elternhaus nach Gemeindetyp

### Individuelle Ressourcen und Lebenslage: Ausbildung und Erwerbsbeteiligung

Aus Anlass verschiedener, nicht eindeutiger Forschungsergebnisse zum Zusammenhang zwischen dem Bildungsniveau der Jugendlichen und dem Auszugsverhalten haben Konietzka und Huinink (2003: 291) vorgeschlagen, statt des erreichten Bildungsniveaus der Jugendlichen die Bildungsbeteiligung und Bildungsübergänge als erklärende Variablen zu verwenden. Entsprechende Modelle ergeben für Westdeutschland, dass insbesondere die Art der 1996 besuchten Schule (Variable *B*) und die Informationen über Beendigungen des Schulbesuchs 1997 bzw. 1998 (Variablengruppe *V*) eng mit dem Auszugsverhalten zusammenhängen. Statistisch signifikant sind ebenfalls die zeitveränderlichen Variablen zum Schulartwechsel 1997 bzw. 1998 (Variablengruppe *W*) (siehe Abb. 4.5-4.7).

Für Ostdeutschland weist zwar keine dieser Variablen einen signifikanten partiellen Erklärungsbeitrag im Vergleich zum vollen Modell mit allen erklärenden Variablen (M2) auf (siehe Tab. 4.4). Berechnet man diese Effekte aber im Vergleich zum gewählten Endmodell 21, werden auch in Ostdeutschland für den Schulbesuch (*B*) und, etwas schwächer, für den Schulbesuchswechsel (Variablengruppe *V*) signifikante Ergebnisse ermittelt (siehe auch Tab. A.2).

Die Modellergebnisse lassen sich wie folgt zusammenfassen: Die Wahrscheinlichkeit auszuziehen steigt im Vergleich zum Besuch einer allgemein bildenden Schule rapide an, wenn Jugendliche eine andere Schule (berufliche Schule, Fachhochschule und Hochschule) oder keine Schule mehr besuchen (siehe Abb. 4.5). Dies gilt insbesondere für Ostdeutschland. In Westdeutschland gibt es jedoch keinen statistisch signifikanten Effekt für Studenten einer Hochschule.

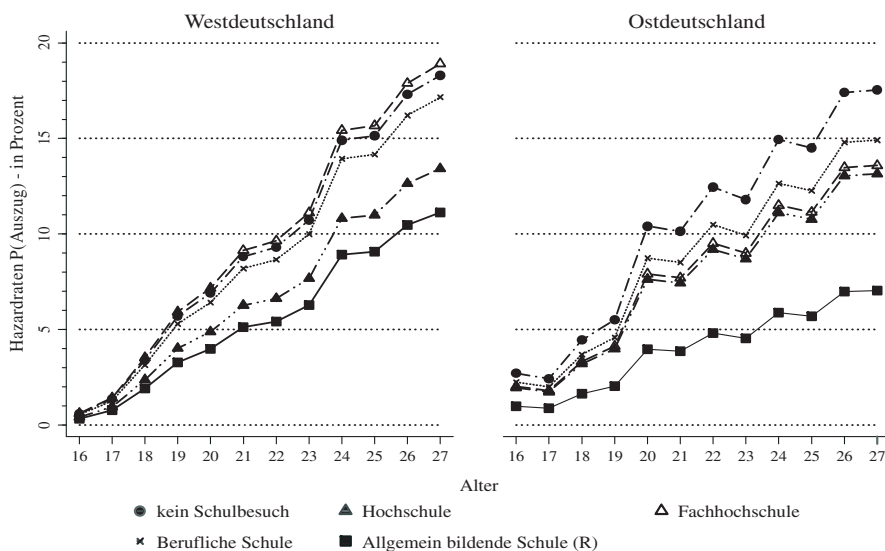


Abbildung 4.5.: Hazardraten zum Auszug aus dem Elternhaus nach dem Schulbesuch (1996)

Abbildung 4.6 zeigt, dass am Ende einer Schulbesuchsepisode Auszüge zunehmen. Des Weiteren sind Wechsel vom Besuch einer allgemein bildenden Schule zu einer beruflichen Schule außer in Ostdeutschland mit einer höheren Auszugsneigung verbunden (siehe Abb. 4.7). Insgesamt sprechen damit die Ergebnisse für einen engen Zusammenhang zwischen Ausbildungsbeginn bzw. -ende und dem Auszug aus dem Elternhaus (vgl. Konietzka und Huinink 2003: 302).

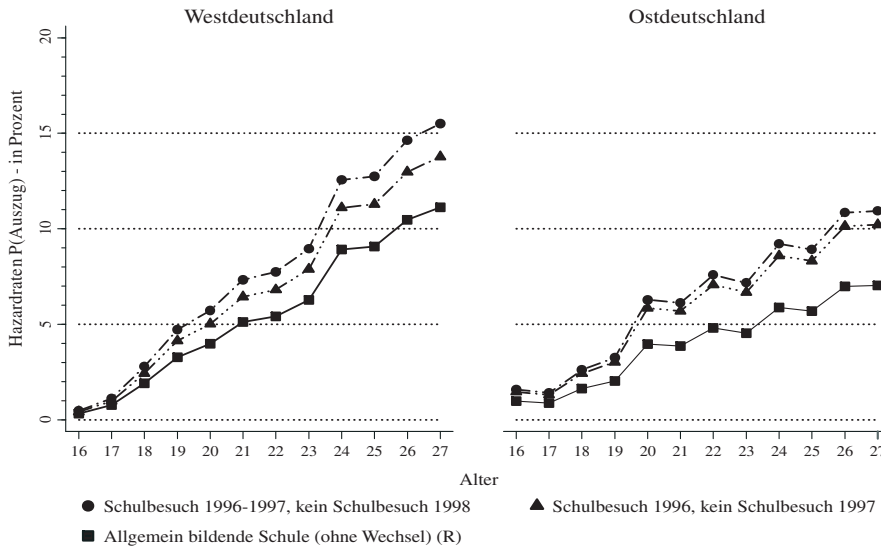


Abbildung 4.6: Hazardraten nach dem Ende des Schulbesuchs

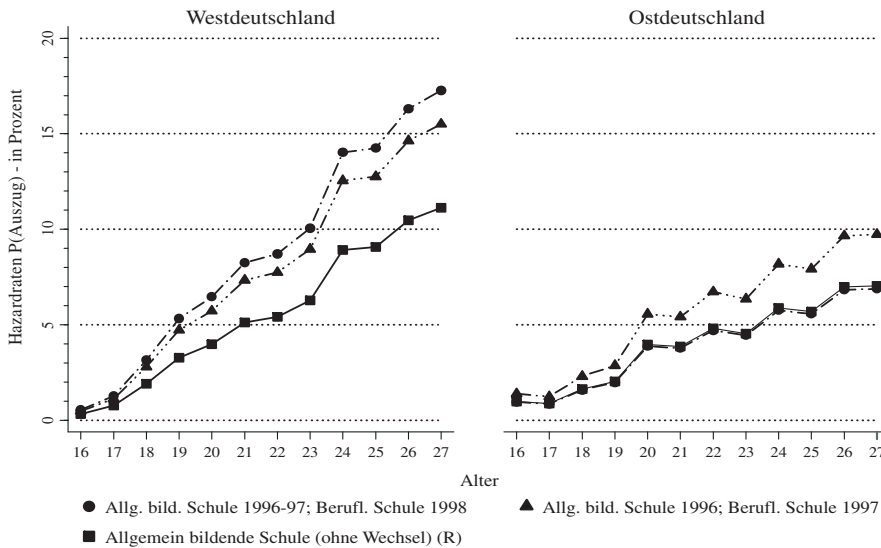


Abbildung 4.7: Hazardraten zum Auszug aus dem Elternhaus nach Schulartwechseln (1996/97; 1996-97/98)

In Bezug auf die Erwerbsbeteiligung ergaben die Modellierungen deutlich geringere Zusammenhänge in Westdeutschland als in Ostdeutschland. Abbildung 4.8 zeigt nur für Angestellte und Beamte in Westdeutschland eine stärkere Auszugsneigung als für nichterwerbstätige Jugendliche (siehe Tabelle A.2). Für Beamte wird auch in Ostdeutschland eine höhere Auszugswahrscheinlichkeit geschätzt. Sonst unterscheiden sich die anderen Kategorien nicht von der Referenzkategorie. Zu erwarten war für erwerbstätige Jugendliche eine höhere Auszugsneigung als bei Nichterwerbstätigen (zumeist Schüler). Dies trifft offensichtlich nur tendenziell zu. Die Ergebnisse geben jedoch keinen Aufschluss über die Gründe der besonderen Auszugswahrscheinlichkeiten von jungen Beamten.

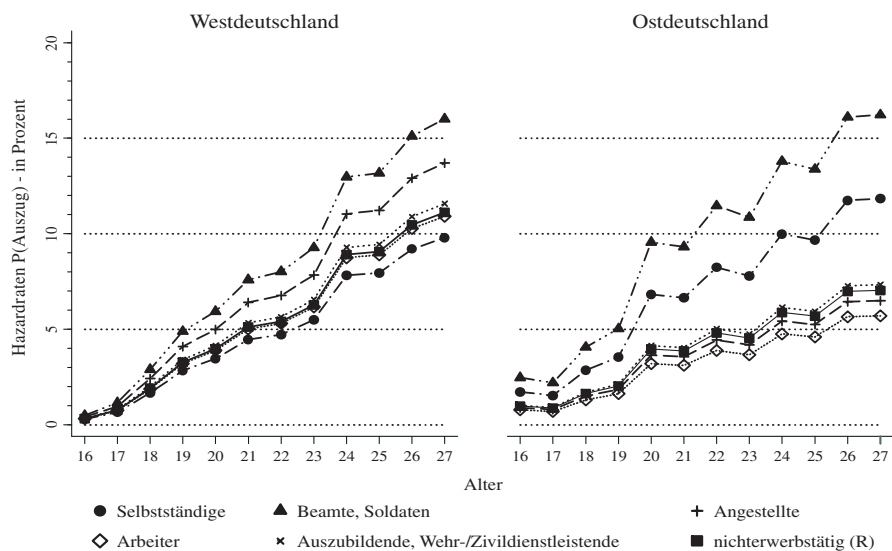


Abbildung 4.8: Hazardraten zum Auszug aus dem Elternhaus nach der Stellung im Beruf (1996)

Ausgehend von der Annahme, dass bei einem Arbeitsplatzwechsel die Auszugswahrscheinlichkeit steigt, wurden Hazardraten für den Betriebswechsel geschätzt. Die in Abbildung 4.9 dargestellten Ergebnisse bestätigen diese Annahme. Ersatzweise bilden hier Auszubildende ohne Betriebswechsel die Referenzgruppe, die sich nicht signifikant von der sonst verwendeten Referenzkategorie unterscheiden. Während aber in Ostdeutschland nur die Hazardrate des Betriebswechsels zwischen 1995 und 1996 signifikant ist, gilt dies in Westdeutschland für Betriebswechsel zwischen 1996 und 1997 sowie 1997 und 1998.

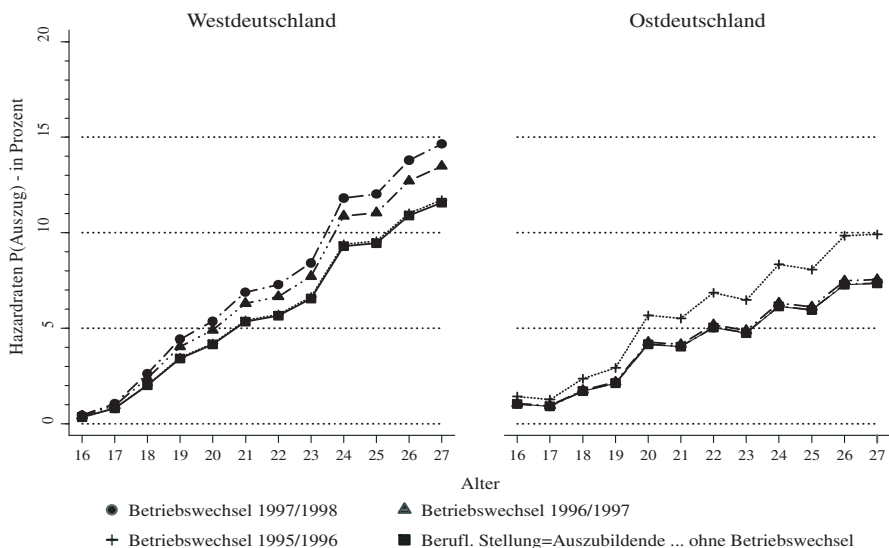


Abbildung 4.9: Hazardraten zum Auszug aus dem Elternhaus nach Betriebswechseln

### Eigenschaften der Familie

Hinsichtlich der Zusammenhänge zwischen dem Auszug aus dem Elternhaus und der sozioökonomischen Lage der Familie hat die Skizze des Forschungsstandes höchstens schwache Zusammenhänge gezeigt (siehe Abschnitt 4.1). Die mit dem Mikrozensuspanel geschätzten Modelle passen zu diesem Tenor, wenn auch im Einzelnen wichtige Unterschiede festzustellen sind. Während sich in Westdeutschland für die Variablen Äquivalenzeinkommen (Variable  $E$ ), Familientyp ( $F$ ), Geschwisterzahl ( $K$ ) und Stellung im Beruf der Familienbezugsperson ( $S^{FB}$ ) im Vergleich zum vollen Modell 2 (siehe Tab. 4.3) signifikante partielle Zusammenhänge ergeben, trifft dies in Ostdeutschland zunächst nur für die Geschwisterzahl, im gewählten Endmodell 21 aber auch für den Familientyp zu.

Betrachtet man das Äquivalenzeinkommen (siehe Abb. 4.10), d. h. das bedarfsgewichtete Pro-Kopf-Haushaltseinkommen, wird für Westdeutschland eine hohe Auszugsneigung, wie erwartet, für die unterste Einkommensklasse geschätzt (siehe Seite 45). Die geringsten Auszugswahrscheinlichkeiten zeigen sich für die mittlere Einkommensklasse. Die der obersten Einkommensklasse ist am größten.

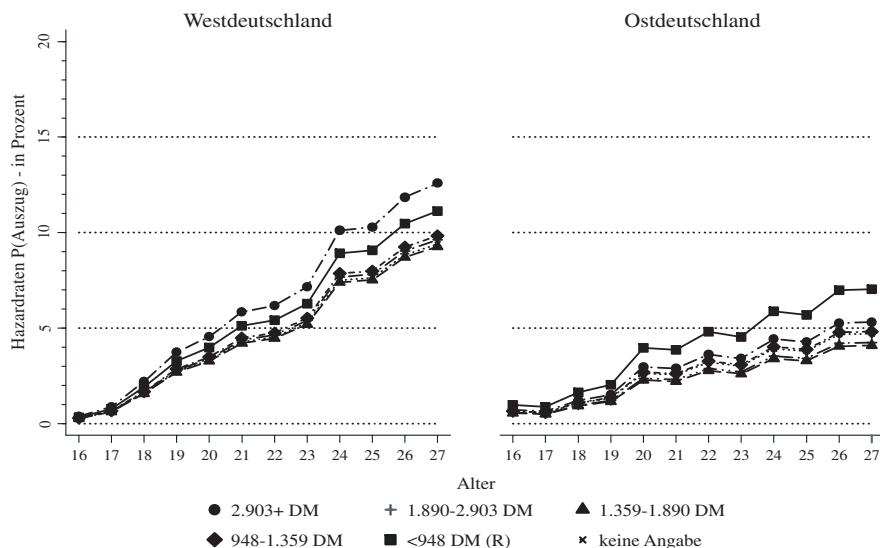


Abbildung 4.10: Hazardraten zum Auszug aus dem Elternhaus nach Äquivalenzeinkommen 1996

Damit deutet sich für Westdeutschland ein leicht U-förmiger Zusammenhang zwischen dem Einkommen und dem Auszug an. Allerdings sind diese Unterschiede bis auf den Koeffizienten für die mittlere Einkommensklasse (1.359-1.890 DM; z-Wert = -1,71) nicht signifikant. Dagegen nimmt in Ostdeutschland mit Ausnahme der obersten Einkommensgruppe die Auszugswahrscheinlichkeit mit zunehmendem Einkommen signifikant ab (siehe Tabelle A.2).

Wie erwartet, zeigen in Abbildung 4.11 die Hazardraten für Jugendliche aus Familien allein erziehender Eltern eine höhere Auszugswahrscheinlichkeit an als für Jugendliche, die mit beiden Eltern zusammenleben. Die Odds-Ratios betragen in beiden Landesteilen rund 1,2 bzw. 1,3.

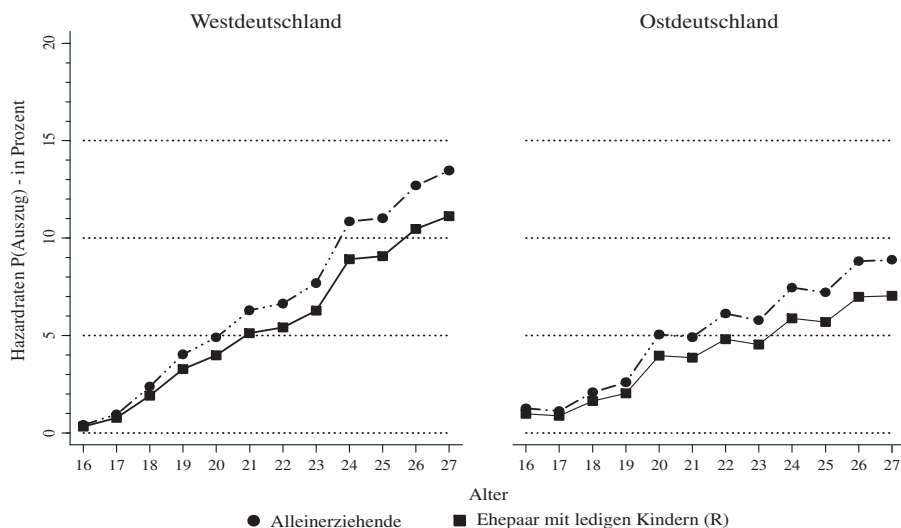


Abbildung 4.11: Hazardraten zum Auszug aus dem Elternhaus nach Familientyp 1996

Abbildung 4.12 zeigt, wie die Auszugswahrscheinlichkeiten mit zunehmender Familiengröße bzw. in Abhängigkeit von der Geschwisterzahl steigen, wobei die Unterschiede zwischen West- und Ostdeutschland auffällig und signifikant sind (siehe Seite 60). Im Gegensatz zu Ostdeutschland unterscheiden sich in Westdeutschland Jugendliche ohne Geschwister oder mit ein bzw. zwei Geschwistern hinsichtlich des Auszugsverhaltens nicht voneinander. Davon verschieden sind Jugendliche mit drei und mehr Geschwistern, allerdings mit einer im Vergleich zu Ostdeutschland niedrigeren Auszugsrate. Dabei ist aber zu beachten, dass diese Gruppe in Ostdeutschland nur rund zwei Prozent der Risikopopulation von 1996 umfasst (siehe Tabelle A.1).

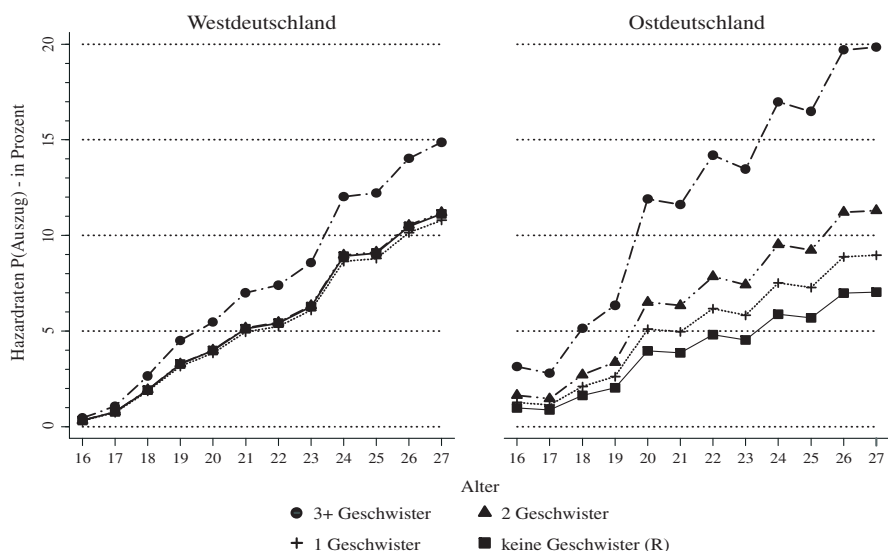


Abbildung 4.12: Hazardraten zum Auszug aus dem Elternhaus nach der Geschwisterzahl 1996

Während in Ostdeutschland keine statistisch signifikanten Unterschiede in Bezug auf die berufliche Stellung der Familienbezugsperson festzustellen sind, werden in Westdeutschland für Jugendliche, deren Familienbezugspersonen (i. d. R. Väter) Selbstständige (inkl. mithelfende Familienangehörige)<sup>26</sup> oder Nichterwerbstätige sind, mit Odds-Ratios von 0,77 bzw. 0,87 signifikant geringere Auszugswahrscheinlichkeiten geschätzt als für die Referenzkategorie der Arbeiter und Angestellten (siehe Abb. 4.13). Für diese beiden westdeutschen Gruppen sind vermutlich unterschiedliche Erklärungen relevant. Davon ausgehend, dass mit dem Äquivalenzeinkommen bereits wichtige Kriterien der sozioökonomischen Lage der Herkunftsfamilien abgebildet werden, sind die Zusammenhänge des Auszugsverhaltens mit der beruflichen Stellung der Familienbezugsperson eher als Residualeffekte aufzufassen. So dürfte bei nichterwerbstätigen Familienbezugspersonen eine im Vergleich zu Erwerbstätigen finanziell prekärere Situation für einen längeren Verbleib der Jugendlichen in der Familie verantwortlich sein. Eine nachträgliche Auswertung zeigt, dass Jugendliche mit einer nichterwerbstätigen Familienbezugsperson in der Einkommensgruppe unter 948 DM mit einem Anteil von rund 21 Prozent stark vertreten sind, während diese Einkommenskategorie durchschnittlich nur rund acht Prozent umfasst. Dagegen ist bei selbstständigen Familienbezugspersonen aufgrund der hohen Anteile von Hauseigentum in dieser Gruppe eher davon auszugehen, dass in diesen Familien ein komfortables Wohnangebot für Jugendliche Anreize bietet, länger im Elternhaus zu wohnen als bei anderen Familien.

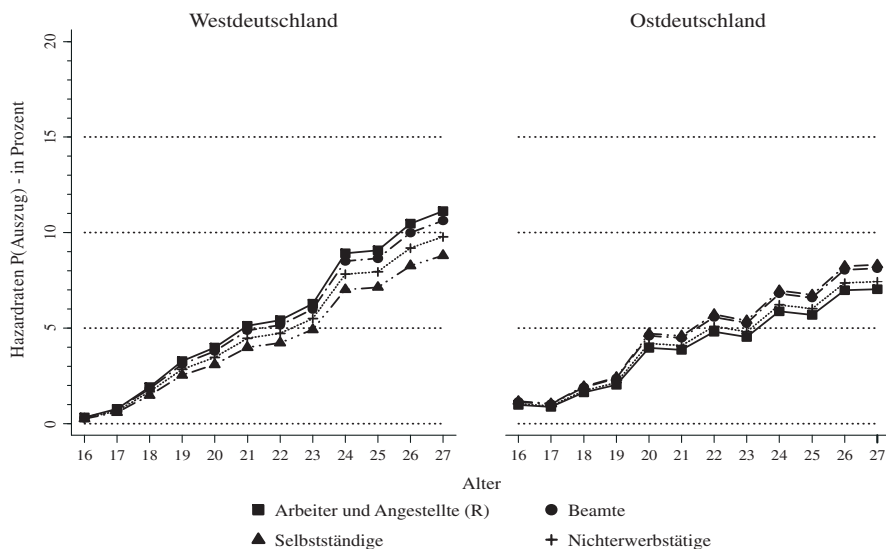


Abbildung 4.13: Hazardraten zum Auszug aus dem Elternhaus nach der Stellung im Beruf der Familienbezugsperson 1996

26 Differenziert man die auf den ersten Blick heterogene Gruppe der Selbstständigen in freie Berufe, Selbstständige in der Landwirtschaft und sonstige Selbstständige, ergeben sich keine statistisch signifikant verschiedenen Hazardraten.

## 4.4 Zusammenfassung

Der Auszug aus dem Elternhaus stellt wohl eines der markantesten Ereignisse dar, mit dem Kinder Unabhängigkeit von den Eltern und Selbstständigkeit erreichen. Bis zum Alter von 25 Jahren haben die meisten Jugendlichen diesen „normalen“ Schritt getan. Auch wenn man für alle Jugendlichen in Bezug auf den Verbleib im Elternhaus bzw. den Auszug vom gleichen Ziel ausgehen kann, sind dennoch unterschiedliche Verhaltensmuster und Gelegenheitsstrukturen zu berücksichtigen, die sich im Zeitpunkt bzw. dem Alter des Auszugs niederschlagen.

Den größten statistischen Erklärungsbeitrag in den Verlaufsanalysen besitzt, wie in allen vergleichbaren nationalen und internationalen Analysen zum Auszug aus dem Elternhaus, die Variable Geschlecht, d. h. Frauen ziehen aufgrund unterschiedlicher Altersnormen der Partnerwahl und Familiengründung früher aus als Männer. Der zweitstärkste Zusammenhang des Auszugsverhaltens zeigt sich für den Gemeindetyp. Je größer und urbaner die Wohngemeinde, desto höher ist die Auszugswahrscheinlichkeit. Dies dürfte mit den zwischen Stadt und Land unterschiedlichen Lebensentwürfen von Jugendlichen und den räumlich verschiedenen Verhältnissen von Mietwohnungen und Hausbesitz zu erklären sein. Die in Ostdeutschland höheren Auszugswahrscheinlichkeiten sind hauptsächlich darauf zurückzuführen, dass im Vergleich zu Westdeutschland Frauen häufiger als Männer ausziehen. Des Weiteren hängt in Ostdeutschland die Geschwisterzahl stärker mit dem Auszug zusammen als in Westdeutschland.

Es war insbesondere zu erwarten, dass nach einem Abschluss der Ausbildung die i. d. R. damit einhergehende Verfügbarkeit über eigene Einkommen die Auszugsneigung fördert. In Bezug auf partielle Effekte ergeben sich jedoch schwächere Zusammenhänge für Faktoren der Bildungs- bzw. Erwerbsbeteiligung und des Einkommens als für die obigen soziodemografischen Merkmale. Diese Zusammenhänge sind zudem in West- und Ostdeutschland etwas unterschiedlich strukturiert. Die Auszugswahrscheinlichkeit ist höher, wenn Jugendliche keine allgemein bildende Schule mehr besuchen. Ausgenommen hiervon ist der Hochschulbesuch in Westdeutschland. Für Erwerbstätige zeigen sich Zusammenhänge zwischen Arbeitsplatzmobilität und räumlicher Mobilität. Die Auszugswahrscheinlichkeiten steigen, falls im Jahr vor der Befragung ein Betriebswechsel stattfand. Bemerkenswerterweise steht die Quelle des Lebensunterhalts der Jugendlichen in keinem statistisch signifikanten Zusammenhang mit dem Auszug aus dem Elternhaus. Die Effekte der Erlangung ökonomischer Unabhängigkeit von den Eltern spiegeln sich vielmehr im Einfluss der Variablen zur Bildungs- und Erwerbsbeteiligung wider.

Wie zu erwarten, neigen Jugendliche aus finanziell schlechter gestellten Haushalten dazu, früher auszugehen. Hinsichtlich der finanziellen Lage der Herkunftsfamilie hängt das hier verwendete Äquivalenzzettoeinkommen mit dem Auszugsverhalten insgesamt betrachtet nur schwach zusammen. Mit Ausnahme der Familiengröße bzw. Geschwisterzahl weisen auch weitere Variablen zur Herkunftsfamilie eine eher untergeordnete Rolle auf. Dies gilt insbesondere für die berufliche Stellung der Familienbezugsperson bzw. des Vaters.

Die Analysen zum Auszug junger Erwachsener aus dem Elternhaus dieses Kapitels behandeln nur Wegzüge. Insofern keine Zielzustände (z. B. bei Heirat oder Studium) erfasst werden, kann die Selektivitätsfrage, ob die Wegzüge bzw. Ausfälle mit dem Statuswechsel selbst verbunden sind, nicht direkt beantwortet werden. Da aber die meisten Wegzüge dieser überdurchschnittlich regional mobilen Teilpopulation Auszüge aus dem Elternhaus darstellen, geben die Auszugsanalysen dennoch wichtige Hinweise zur Klärung der Selektivitätsfrage.

Unter Berücksichtigung dieser Einschränkung lassen sich deshalb nur vorläufige Schlussfolgerungen in Bezug auf die Fragen zur Stichprobenselektivität ziehen. Die statistisch schwachen Zusammenhänge zwischen der sozioökonomischen Lage des Elternhauses mit dem Auszug las-



sen vermuten, dass mit diesen Variablen verbundene potenzielle Verzerrungen bei Analysen räumlich mobiler Personen des Mikrozensuspanels eher gering sein werden. Der starke Zusammenhang zwischen Ausfall bzw. Auszug und Alter sowie den Variablen Geschlecht und Gemeindetyp, für die in den Modellen zeitkonstante Effekte angenommen wurden, weist darauf hin, dass Ausfallkorrekturen mithilfe entsprechend konstruierter Faktoren möglich sind, mit denen mit dem Kehrwert der Wahrscheinlichkeit räumlicher Immobilität gewichtet wird.

Zwar ist insgesamt betrachtet der Einfluss zeitveränderlicher Variablen auf die Auszugsneigung als moderat zu bezeichnen (siehe Tab. 4.4), doch hinsichtlich der Befunde, dass Ausbildungsperioden nach der allgemein bildenden Schule sowie Betriebswechsel zum Anstieg der Auszugswahrscheinlichkeiten führen, ist zumindest teilweise zu erwarten, dass Ausfälle und Statuswechsel zusammenhängen. Jedoch deuten die für Hochschulbesucher in Westdeutschland geschätzten Auszugswahrscheinlichkeiten, die sich nicht signifikant von den Hazardraten für Besucher allgemein bildender Schulen unterscheiden, auch darauf hin, dass Analysen dieser Übergänge und Teilpopulationen auf Basis räumlich mobiler Personen nicht wesentlich verzerrt sein dürften. Ob und in welchem Maße Statuswechsel zum Auszug aus dem Elternhaus bzw. allgemein dem Wegzug aus dem Haushalt des Auswahlbezirks führen, ist das Kernthema der folgenden Untersuchungen.

## 5 Vom Besuch der gymnasialen Oberstufe bis zum Abitur

Die bildungspolitische Diskussion um die Ergebnisse der Pisa-Untersuchung zu der im internationalen Vergleich in Deutschland hohen sozialen Selektivität der Bildungsbeteiligung im Sekundarbereich (Baumert et al. 2001) hat wieder darauf aufmerksam gemacht, dass in der amtlichen Bildungsstatistik Informationen über sozioökonomische Merkmale des Elternhauses der Schüler vollständig fehlen und außerdem den dringenden Bedarf an längsschnittlichen Daten aufgezeigt (Avenarius et al. 2003; Kristen et al. 2005; Schmidt und Weishaupt 2004; Weishaupt und Fickermann 2001; Weißhuhn 2001).

Um Bildungsübergänge im sozialen und regionalen Kontext untersuchen zu können, werden umfangreiche Datenbasen benötigt. In dieser Hinsicht bietet der Mikrozensus aufgrund seines enormen Stichprobenumfangs und der infolge der Auskunftspflicht sehr hohen Ausschöpfung für die Bildungs- und Sozialstrukturforschung einige Vorteile. Darüber hinaus können wegen seiner Durchführung als Haushaltsstichprobe Merkmale der Ausbildungsabsolventen mit Angaben der sozialen Stellung der Familie zusammengeführt werden. Während mit der Schulstatistik die Selektivität in der Sekundarstufe II nur auf aggregierter Ebene und ohne Differenzierung nach sozialer Herkunft der Schüler betrachtet werden kann, wäre es ein großer Vorteil, wenn mit dem Mikrozensuspanel entsprechende Quoten auf Individualdatenbasis geschätzt werden könnten. Aufgrund der obigen Ergebnisse zum Auszug aus dem Elternhaus ist bei diesen Analysen eine geringe und vermutlich korrigierbare Verzerrung durch räumliche Mobilität zu erwarten.

Dieses Kapitel behandelt die Analysemöglichkeiten von Bildungsverläufen am Beispiel des Besuchs der gymnasialen Oberstufe bis zum Abitur und geht dabei insbesondere auf die Ausfallproblematik ein. Diese auf den ersten Blick vielleicht einfach erscheinende Fragestellung stellt nicht nur hinsichtlich der fehlenden Weiterverfolgung wegziehender Personen besondere Anforderungen. Im weiteren Sinne hängen die Analysemöglichkeiten auch davon ab, wie gut die zu einem Panel zusammengeführten Querschnittsangaben im Längsschnitt auswertbar sind. Nach dem Konzept der Berichtswoche, das auf die Erhebung von Querschnittsergebnissen ausgerichtet ist, werden Angaben zum Schulbesuch und Bildungsabschluss im Mikrozensus jährlich im Frühjahr erfragt. Da aber der erfolgreiche Abschluss der gymnasialen Oberstufe erst später zum Ende des Schuljahres feststeht, kann erst in der Befragung ein Jahr darauf ermittelt werden, welche Schüler das Abitur erreicht haben.

Im Folgenden werden zunächst die in den Daten vorliegenden Ausgangsinformationen und die Vorgehensweise bei der Ermittlung des Abschlusses beschrieben. Im zweiten Abschnitt werden erste Ergebnisse zum Abschluss der gymnasialen Oberstufe mit dem Abitur vorgestellt. Deskriptive Vergleiche der Schüler- und Abiturientenzahlen der Bildungsstatistik und des Mikrozensuspanels stehen im Zentrum des dritten Abschnittes. Darin werden Fragen zur Datenqualität, insbesondere zur Konsistenz der Bildungsangaben aufgegriffen, die mit den Paneldaten erstmals untersuchbar sind. Anschließend folgen nach einer Skizze theoretischer Ansätze zur Erklärung der sozialen Ungleichheiten im Bildungssystem Ergebnisse multivariater Analysen zum Ausfall bzw. Wegzug im Mikrozensuspanel. Das Kapitel schließt mit einer Zusammenfassung der wichtigsten Ergebnisse.

## 5.1 Ausgangsinformationen und Vorgehensweise bei der Ermittlung des Abiturabschlusses

Der Besuch der gymnasialen Oberstufe (Klassenstufe 11-13; Sekundarstufe II) folgt in der Regel im Anschluss an die zehnte Klasse (Sekundarstufe I). In Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen kann das Abitur bereits nach zwölf Schuljahren erreicht werden.<sup>27</sup> Die gymnasiale Oberstufe umfasst in diesen Fällen die Klassenstufen 10-12 oder 11-12. Den Kern der gymnasialen Oberstufe bilden das Gymnasium, die Integrierte Gesamtschule sowie Freie Waldorfschulen. In manchen Bundesländern gehören aber auch berufliche Gymnasien oder Fachgymnasien dazu. Die Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) fasst deshalb unter der gymnasialen Oberstufe verschiedene Schularten zusammen (siehe Abb. 5.1).

Als reguläre Abschlüsse werden die Allgemeine Hochschulreife (Abitur) und die Fachgebundene Hochschulreife vergeben. In einzelnen Bundesländern wird bei vorliegendem Versetzungszugnis in die 13. (bzw. 12.) Klasse der Abschluss der Fachhochschulreife vergeben, die eigentlich ein beruflicher Bildungsabschluss ist. Neben der Fachhochschulreife können sowohl die Fachgebundene Hochschulreife als auch die Allgemeine Hochschulreife auch an beruflichen Schulen erlangt werden (KMK 2003). Diese länderspezifischen Unterschiede im Rahmen von Mehrthemenumfragen zu erfassen, stellt eine große Herausforderung dar.

Im Mikrozensus stehen für die Ermittlung des Bildungsverlaufs in der gymnasialen Oberstufe bis zum Abitur jährlich Angaben zum Schulbesuch und zum erreichten Bildungsabschluss zur Verfügung. Die Frage, ob man gegenwärtig eine Schule besucht, wird allen Befragten gestellt. Falls diese bejaht wird, geben die Befragten im Anschluss daran die Art der besuchten Schule an. Besucher einer allgemein bildenden Schule sollen die besuchte Klassenstufe (1-4, 5-10, 11-13) ankreuzen. Die jeweilige Schulart (Hauptschule, Realschule, Gymnasium usw.) wird seit 1991 nicht mehr erfragt. Der Besuch einer beruflichen Schule, der Fachhochschule und der Hochschule wird mit extra Kategorien erfasst.

Im Unterschied zur Darstellung der Bildungsstruktur durch die KMK in Abbildung 5.1 gehören aber im Mikrozensus Besucher beruflicher Gymnasien und Fachgymnasien zu beruflichen Schulen. In Bezug auf solche Abgrenzungsfragen ist darauf hinzuweisen, dass Befragte, die sich für eine schriftliche Beantwortung entschieden haben, im Fragebogen keine weiteren Erläuterungen zu den abgefragten Begriffen erhalten. Immerhin rund 15 Prozent aller Angaben stammen aus schriftlichen Auskünften. Insbesondere diese Befragten werden deshalb vermutlich die Fragen auf Basis ihrer Alltagsdefinition beantworten.

Lediglich im Interviewerhandbuch werden die zur gymnasialen Oberstufe gehörenden Schularten aufgeführt: Gymnasien, Integrierte Gesamtschulen, Freie Waldorfschulen, Sonderschulen,<sup>28</sup> Abendgymnasien und Kollegs. Laut Interviewerhandbuch zählen die folgenden Schulen, an denen ebenfalls die Allgemeine oder Fachgebundene Hochschulreife und die Fachhochschulreife erworben werden können, zu den beruflichen Schulen: Fachoberschulen, Fachgymnasien / Berufliche Gymnasien und Technische Oberschulen.

Im Mikrozensus richten sich Fragen zum allgemeinen Schulabschluss nur an Personen im Alter von 15 Jahren und älter, die keine allgemein bildende Schule besuchen. An eine Unterstichprobe mit variablem Auswahlatz von durchschnittlich 0,45 Prozent der Haushalte wird im Rah-

27 Die Beschreibung bezieht sich auf den Zeitraum bis Ende der 1990er Jahre. Mittlerweile haben eine Reihe von Ländern Reformen zur Verkürzung auf eine zwölfjährige Dauer bis zum Abitur begonnen.

28 Sonderschulen zählen laut KMK nicht zur gymnasialen Oberstufe. An einigen Sonderschulen kann auch das Abitur erworben werden; die Zahl von Sonderschülern in den Klassen 11-13 ist sehr gering.



men der in den Mikrozensus integrierten EU-Arbeitskräfteerhebung die Frage gestellt, an welcher Schulform der Abschluss erreicht wurde. Der Zeitpunkt des Abschlusses wird erst ab 1999 im Rahmen der Unterstichprobe erfragt, jedoch ist diese Information nicht im Panelfile enthalten (siehe Abschnitt 2.1). Für Besucher allgemein bildender Schulen liegen somit keine Angaben über den zum Befragungszeitpunkt bereits erreichten Schulabschluss vor.

Während die Mikrozensusenerhebung in der Regel im April durchgeführt wird, liegt das Abitur teilweise erst im Juni oder Juli, am Ende des Schuljahres, vor.<sup>29</sup> Aus diesen Gründen kann ein erfolgreicher Bildungsweg nur im Vergleich von aufeinander folgenden Befragungen ermittelt werden. Dabei wird wie folgt verfahren: Ist zu einem Zeitpunkt (z. B. 1996) der Besuch der gymnasialen Oberstufe angegeben, im Folgejahr (z. B. 1997) jedoch nicht mehr, und wird zu diesem Zeitpunkt erstmals der Abschluss Fachhochschulreife bzw. Allgemeine oder Fachgebundene Hochschulreife genannt, kann davon ausgegangen werden, dass dieser Abschluss zum Ende des letzten Schuljahres (z. B. 1996) erfolgt ist. Hierzu ein weiteres Beispiel: Für Schüler, die 1996 die elfte Klasse besuchen und bei einem regulären Schulverlauf 1998 die Oberstufe mit dem Abitur abschließen, kann dieses Ereignis erst im Mikrozensus 1999 festgestellt werden.

Bei Schülern mit Klassenwiederholungen sind Untererfassungen des Abschlusses zu erwarten, da die Beobachtungszeit des Mikrozensuspanels auf maximal vier Jahre begrenzt ist. Unter Kontrolle des Alters sollte jedoch eine Absolventenquote näherungsweise berechnet werden können. Dieses Vorgehen setzt allerdings voraus, dass die Absolventen bis dahin nicht fortgezogen sind. Sofern die Abiturienten bei einem Umzug, beispielsweise an den Hochschulort, noch ihren früheren Wohnsitz als Haupt- oder Nebenwohnsitz beibehalten, liegen jedoch Angaben vor.

## 5.2 Deskriptive Ergebnisse zum Besuch der gymnasialen Oberstufe und dem Abschluss mit dem Abitur

Zur Ermittlung der Abiturienten werden aus dem Mikrozensuspanel alle erfolgreich zusammengeführten Personen ausgewählt, die einen Besuch der gymnasialen Oberstufe in einem der Erhebungszeitpunkte 1996 bis 1999 angaben. Im Folgenden wird der sprachlichen Einfachheit halber zum Abitur (Allgemeine und Fachgebundene Hochschulreife) auch die Fachhochschulreife gerechnet. Um Vergleiche mit den nur für Querschnitte vorliegenden Daten der Bildungsstatistik zu ermöglichen, werden zunächst alle Personen berücksichtigt, die an mindestens einer Befragung teilgenommen haben. Tabelle 5.1 enthält die ungewichteten Fallzahlen der so ermittelten Bestände und Veränderungen zum Besuch der gymnasialen Oberstufe von 1996 bis 1999. Auffällig ist die hohe Mobilität. Die Zu- und Abgänge summieren sich auf jeweils rund 50 Prozent des entsprechenden Anfangsbestandes.

Unter den Zugängen aus anderen Schulformen befinden sich überwiegend Schüler der Sekundarstufe I. Zu erwarten wäre, dass diese Zugänge aus der Sekundarstufe I etwa dem Umfang einer der drei Klassenstufen 11-13 entsprechen, also rund einem Drittel (bzw. der Hälfte bei zwölfjährigem Regelschulbesuch). Im Mikrozensuspanel betragen diese Zugänge aber etwa 40 bis 45 Prozent und liegen damit deutlich höher.

29 Diese Angaben beziehen sich auf das Fragenprogramm der Querschnittserhebungen bis 2004. Seit 2005 wird der Mikrozensus unterjährig durchgeführt, sodass für das ganze Jahr Angaben vorliegen. In diesem Zusammenhang sind unter anderem die Fragen zum Besuch von Hochschulen und Schulen geändert worden. Erfragt wird nun der Schulbesuch in den letzten vier Wochen sowie in den letzten zwölf Monaten. Darüber hinaus geben die Befragten an, in welchen Jahren sie ihren höchsten allgemeinen und beruflichen Abschluss erreicht haben.

Tabelle 5.1: Besucher der gymnasialen Oberstufe 1996-1999

Bestand und Bestandsentwicklung	Insg. n	in %	West n	in %	Ost n	in %
Besucher der gymnasialen Oberstufe 1996	1.609	100	1.378	100	231	100
- Abgänge						
Abitur April 1997	313	19,5	261	18,9	52	22,5
Wechsel in andere Schulen etc. 1996/97	303	18,8	264	19,2	39	16,9
Ausfälle 1996/97	147	9,1	125	9,1	22	9,5
+ Zugänge						
Wechsel aus anderen Schulen 1996/97	717		611		106	
Stichprobenzugang; 1996 nicht Befragte	125		111		14	
= Besucher der gymnasialen Oberstufe 1997	1.688	100	1.450	100	238	100
- Abgänge						
Abitur April 1998	333	19,7	273	18,8	60	25,2
Wechsel in andere Schulen etc. 1997/98	284	16,8	244	16,8	40	16,8
Ausfälle 1997/98	161	9,5	136	9,4	25	10,5
+ Zugänge						
Wechsel aus anderen Schulen 1997/98	692		597		95	
Stichprobenzugang; 1997 nicht Befragte	127		118		9	
= Besucher der gymnasialen Oberstufe 1998	1.729	100	1.512	100	217	100
- Abgänge						
Abitur April 1999	332	19,2	285	18,8	47	21,7
Wechsel in andere Schulen etc. 1998/99	296	17,1	262	17,4	34	15,7
Ausfälle 1998/99	177	10,2	150	9,9	27	12,4
+ Zugänge						
Wechsel aus anderen Schulen 1998/99	704		607		97	
Stichprobenzugang; 1998 nicht Befragte	130		117		13	
= Besucher der gymnasialen Oberstufe 1999	1.758		1.539		219	

West: Bundesländer mit 13-jährigem Schulbesuch (Westdeutschland einschl. Berlin und Brandenburg)

Ost: Bundesländer mit zwölfjährigem Schulbesuch (Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen)

„Abitur“ umfasst hier neben dem Abschluss der Allgemeinen und Fachgebundenen Hochschulreife auch die Fachhochschulreife.

Quelle: Mikrozensuspanel 1996-1999; eigene Berechnungen (ungewichtete Fallzahlen).

Hinsichtlich der Abgänge ist festzustellen, dass rund 17 bis 19 Prozent der Schüler eines Jahrgangs die gymnasiale Oberstufe verlassen und zu anderen Schulen wechseln oder keine Schule mehr besuchen. Betrachtet man diese Gruppe am Beispiel des Übergangs 1996/97, zeigt sich folgendes Muster: Von den Personen, die 1996/97 die gymnasiale Oberstufe verlassen haben (n = 303), gehen 1997 die meisten Schüler in eine berufliche Schule (41,9 %), gefolgt von Wechseln in

die Sekundarstufe I einer allgemein bildenden Schule (26,7 %),<sup>30</sup> Personen ohne Schulbesuch (23,1 %) und Besuchern der Fachhochschule oder Hochschule (8,3 %).

Im Allgemeinen würde man annehmen, dass die Schüler, die bis zur elften Klasse der Oberstufe aufgerückt sind, überwiegend bis zum Abitur dort verbleiben. Deshalb erscheint der Anteil von Übergängen in berufliche Schulen überraschend hoch. Bei der Teilgruppe der Fachhochschul- und Hochschulbesucher (8,3 %) wird kein Abschluss mit dem Abitur festgestellt. Dies ist darauf zurückzuführen, dass für diese Gruppe 1997 keine entsprechenden Abschlussangaben vorliegen.

Für jeweils nur knapp 20 Prozent der Besucher der gymnasialen Oberstufe eines Erhebungszeitpunkts bzw. Übergangs konnte der Abschluss mit dem Abitur ermittelt werden. Nimmt man die Fachhochschul- und Hochschulbesucher ohne Angabe eines Abschlusses (s. o.) hinzu, erhöht sich der Anteil der Absolventen mit Abitur geringfügig (z. B. 1996/97 von 19,5 % auf 21,0 %). Erst wenn man annimmt, dass alle infolge räumlicher Mobilität ausgefallenen Schulbesucher (rd. 10 %) die gymnasiale Oberstufe mit dem Abitur abgeschlossen haben, wird mit jeweils etwa 30 Prozent insgesamt ein Wert erreicht, der unter Berücksichtigung einer durchschnittlichen Wiederholerquote in der Sekundarstufe II von circa drei Prozent (Cortina und Trommer 2003: 369; siehe hierzu auch Schimpl-Neimanns 2005: 59) näherungsweise dem Anteil einer Klassenstufe entspricht.

Die obigen Salden und ihre Veränderungen haben auf eine ausgeprägte Mobilität des Schulbesuchs aufmerksam gemacht. Dies kann einerseits mit untypischen Bildungskarrieren (Cortina und Trommer 2003) zusammenhängen, andererseits sind aber auch Antwortfehler bzw. inkonsistente Angaben nicht auszuschließen. Nicht zuletzt ist aufgrund der durch räumliche Mobilität verursachten Ausfälle mit einer Untererfassung von Abiturienten zu rechnen.

Für die Beurteilung der Validität der mit dem Mikrozensuspanel ermittelten Abiturientenzahlen werden externe Kriterien benötigt. Wegen fehlender Verlaufsstatistiken auf Individualdatenbasis liegen allerdings keine Erfolgsquoten zur Einschätzung der Selektivität in der Sekundarstufe II vor (Bellenberg 1999).<sup>31</sup> Insbesondere ist unbekannt, wie viele Schüler der elften Klasse das Abitur erreichen oder die Schule ohne Abitur verlassen. Somit fehlen belastbare Vergleichswerte, die geeignet sind, die mit dem Mikrozensuspanel ermittelten Ergebnisse zu überprüfen. Man kennt aber aus der amtlichen Bildungsstatistik die Gesamtzahlen der Schüler in den Klassen 11-13 und der Entlassenen nach Abschlussart. Behelfsweise lassen sich Quasi-Erfolgsquoten berechnen, indem man die Zahl der Absolventen ins Verhältnis zur Zahl der Besucher der elften Klasse zwei Jahre zuvor setzt, bzw. ein Jahr zuvor in den Ländern mit zwölfjähriger Regelschulzeit. Die so geschätzten Erfolgsquoten betragen in den Schuljahren 1995/96 bis 1998/99 für West- bzw. Ostdeutschland 86 bzw. 90 Prozent (Schimpl-Neimanns 2005: 59). Da bei dieser Betrachtung auf Aggregatebene individuelle Veränderungen, wie beispielsweise Schulwechsel, Klassenwiederholungen und Abbruch etc., nicht berücksichtigt werden, wird die Erfolgsquote

30 Dieser mit rund 27 Prozent recht hohe Anteil von Schülern, die in die Sekundarstufe I wechseln, verweist im Zusammenhang mit weiteren im Längsschnitt zu beobachtenden Wechseln auf eine wenig trennscharfe Unterscheidung zwischen den Klassenstufen 5-10 und 11-13.

31 Im Sozio-oekonomischen Panel (SOEP) wird die besuchte Klassenstufe nicht erfragt, sodass auch damit keine Übergangs- und Erfolgsquoten für die gymnasiale Oberstufe berechnet werden können. Schneider (2005: 173) konnte auf Basis des SOEP für 164 von 843 Jugendlichen, die in einem Jahr das Gymnasium besucht haben und für die wenigstens eine gültige Angabe im Folgejahr vorliegt, den Bildungsweg vom letzten Grundschuljahr bis zum Erwerb der allgemeinen Hochschulreife verfolgen. Nach seinen Verlaufsauswertungen besuchen sechs Jahre nach Eintritt in das Gymnasium noch 78 Prozent der Schüler diese Schulform (Schneider 2005: 180). Die Ergebnisse beschränken sich im Wesentlichen auf die Sekundarstufe I.



tendenziell überschätzt. Die Kennziffern markieren gewissermaßen nur die Obergrenze des schulischen Erfolgs in der gymnasialen Oberstufe.

Nach Ergebnissen einer älteren Verlaufsuntersuchung (Pötzsch 1982: 130) erreichten von den Besuchern der elften Klasse der Oberstufe insgesamt 69 Prozent nach drei Jahren und 75 Prozent nach vier Jahren das Abitur. Für Schüler, die schon in der Sekundarstufe I das Gymnasium besucht haben, berichtet Pötzsch eine Erfolgswahrscheinlichkeit von 84 Prozent (nach vier Jahren: 88 %). Orientiert man sich mangels anderer Daten an den Ergebnissen dieser speziellen Studie, können diese Quoten zwischen rund 70 und 90 Prozent als grobe untere Richtwerte für die mit dem Mikrozensuspanel ermittelten Ergebnisse dienen.<sup>32</sup>

Um den oben aufgeworfenen Fragen zu den Problemen bei der Erfassung des Abiturabschlusses aufgrund von Wegzügen und der hohen Mobilität des Schulbesuchs genauer nachzugehen, ist eine altersspezifische Darstellung sinnvoll. Im Folgenden wird in einem ersten Schritt gezeigt, wie viele der Personen, die 1996 die gymnasiale Oberstufe besuchten, bis 1999 das Abitur erreicht haben, bzw. welchen Status sie 1999 einnehmen. Danach folgen Informationen zu den altersspezifischen Übergängen in den dazwischen liegenden Jahren. Aufgrund der in vier Bundesländern kürzeren Regelschulzeit bis zum Abitur werden die Ländergruppen getrennt dargestellt. Die folgenden Analysen beschränken sich auf die für die gymnasiale Oberstufe typische Altersgruppe der beim Start des Panels (1996) 16- bis 19-Jährigen.<sup>33</sup> Die Mehrheit der 1996 17-Jährigen dürfte damals die elfte Klasse besucht und bei regulärem Schulverlauf, d. h. ohne Klassenwiederholung, am Ende der 13. Klasse 1998 das Abitur erreicht haben. Die 1996 17-Jährigen bilden deshalb die Referenzgruppe.

Abbildung 5.2 zeigt, dass in den Ländern mit 13-jähriger Regelschulzeit nur für die Hälfte aller 17-Jährigen bis 1999 das Abitur ermittelt werden kann. Rechnet man die Besucher von Fachhochschulen und Hochschulen hinzu, liegt der Anteil der Abiturienten bei 53 Prozent. Selbst bei den 1996 19-Jährigen erreicht die Abiturientenquote nur knapp 70 Prozent (inkl. Fachhochschul- und Hochschulbesuch). Bei den 16-Jährigen beträgt sie sogar nur 20 Prozent. In dieser Altersgruppe haben 1999 noch rund 36 Prozent das Abitur vor sich.<sup>34</sup> Rund 20 Prozent der jeweiligen Altersgruppe ziehen um, bevor ein Abschluss ermittelt werden kann.<sup>35</sup> Die insgesamt geringe „Erfolgsquote“ könnte auf die Nichterfassung weggezogener Abiturienten zurückzuführen sein.

---

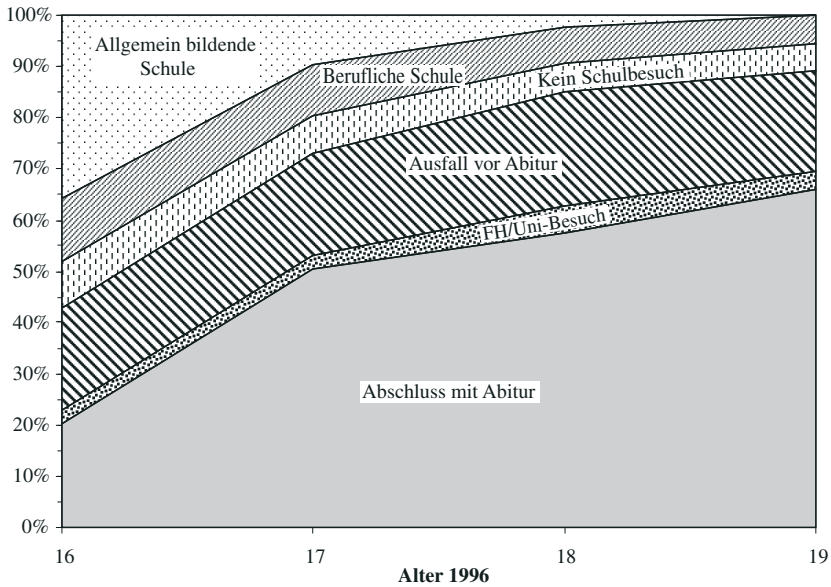
32 Die Studie konzentrierte sich auf eine Stichprobe von 618 Schüler aus zehn Gymnasien in Nordrhein-Westfalen, die außer Gymnasiasten der zehnten Klasse auch Haupt- und Realschüler in die elfte Klasse aufgenommen haben. Der Zeitraum erstreckte sich von September 1977 bis Juli 1980.

33 In den westdeutschen Ländern einschl. Berlin und Brandenburg repräsentieren die 16- bis 19-Jährigen 83 Prozent aller Schüler der gymnasialen Oberstufe. In den anderen vier Bundesländern liegt ihr Anteil bei 90 Prozent. Nicht berücksichtigt werden im Folgenden 15 Personen mit temporären Ausfällen.

34 Dabei handelt es sich überwiegend um Besucher der gymnasialen Oberstufe. In der Kategorie sind auch wenige Personen enthalten, die angeben, die Klassenstufen 5-10 allgemein bildender Schulen zu besuchen.

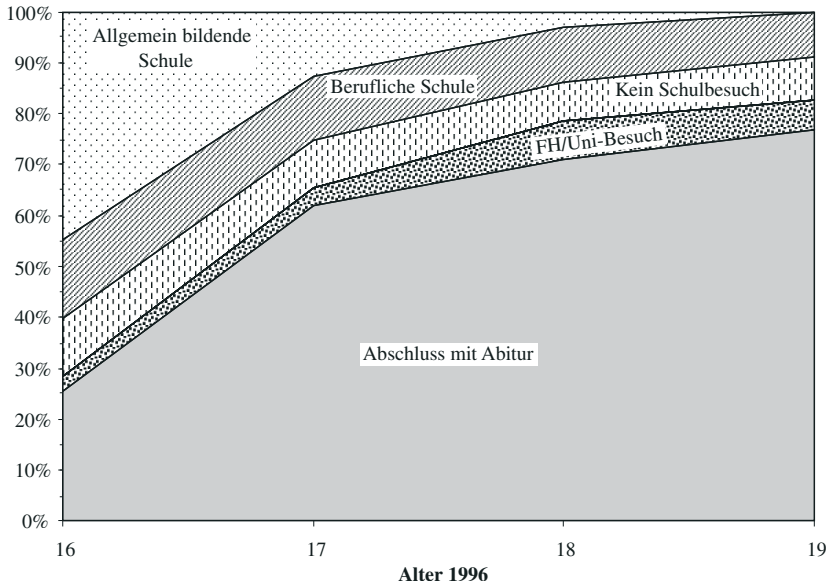
35 Die Gruppe „Abschluss mit Abitur“ enthält auch Personen, die erst nach Ermittlung des Abiturs weggezogen sind. Insgesamt beträgt die Ausfallquote für die 16- bis 19-Jährigen im Westen 28 Prozent, im Osten 37 Prozent.





Quelle: Mikrozensuspanel 1996-1999 (n = 1.129); eigene Berechnungen.

Abbildung 5.2: Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 – Bundesländer mit 13-jährigem Schulbesuch (Insgesamt)

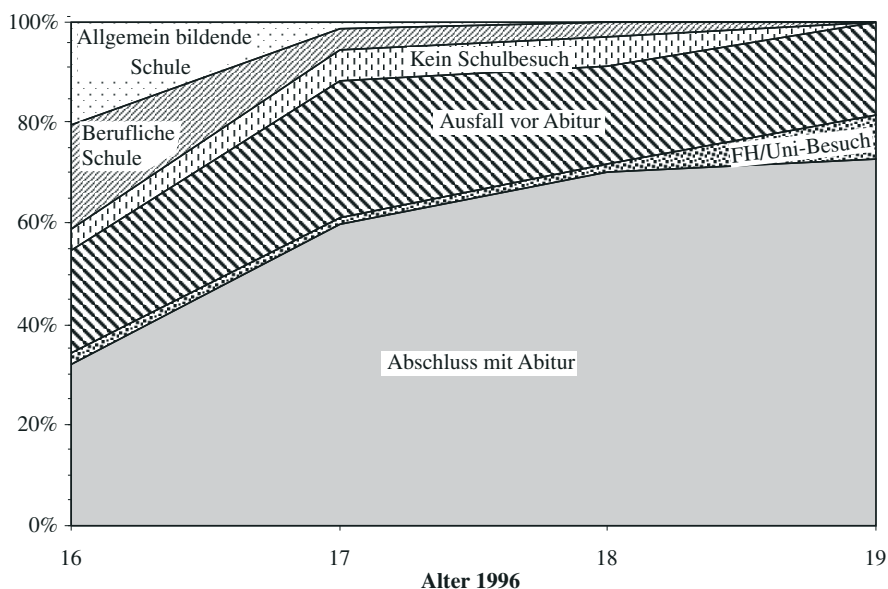


Quelle: Mikrozensuspanel 1996-1999 (n = 821); eigene Berechnungen.

Abbildung 5.3: Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996, räumlich Immobile – Bundesländer mit 13-jährigem Schulbesuch

Wie jedoch die Analyse für die räumlich Immobilen zeigt (siehe Abb. 5.3), erreichen 62 Prozent der 17-Jährigen und 77 Prozent der 19-Jährigen bis 1999 das Abitur. Die Anteile der räumlich immobilen Absolventen mit Abitur sind zwischen sechs und 16 Prozentpunkte höher als die Anteile bei allen Schülern (einschl. räumlich Mobile). Bei den 17- bis 19-Jährigen beträgt die Differenz rund elf bis 13 Prozentpunkte. Nimmt man an, dass sich die Erfolgswahrscheinlichkeiten eines Abiturs zwischen räumlich mobilen und immobilen Schülern nicht unterscheiden, ist zu vermuten, dass von den insgesamt rund 20 Prozent räumlich mobilen Schülern, die vor Ermittlung des Abiturs weggezogen sind, etwas mehr als die Hälfte bis 1999 das Abitur erreicht hat.

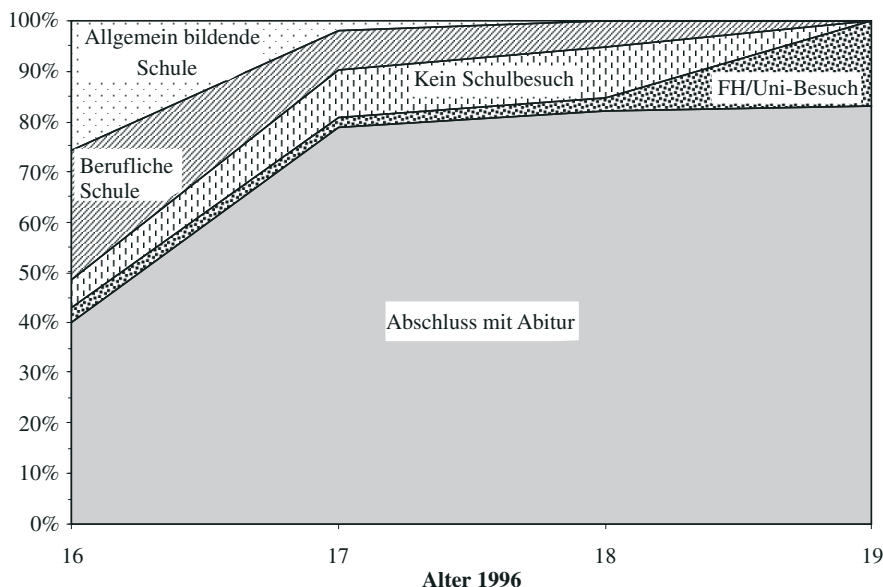
In den vier Bundesländern mit zwölfjährigem Schulbesuch bis zum Abitur sind aufgrund der um ein Jahr längeren Beobachtungszeit im Mikrozensuspanel höhere Anteile der Absolventen mit Abitur zu erwarten. Entsprechend beträgt die „Erfolgsquote“ der 17-Jährigen dort 60 Prozent (siehe Abb. 5.4). Sie liegt damit – trotz einer höheren Ausfallrate – um neun Prozentpunkte höher als in Westdeutschland. Weiterhin sind im Jahre 1999 deutlich geringere Anteile des Besuchs der gymnasialen Oberstufe als in den Bundesländern mit 13-jährigem Schulbesuch festzustellen.



Quelle: Mikrozensuspanel 1996-1999 (n = 209); eigene Berechnungen.

Abbildung 5.4: Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 – Bundesländer mit zwölfjährigem Schulbesuch (Insgesamt)

Betrachtet man die Abiturientenanteile der räumlich Immobilen, lässt sich einerseits im Vergleich zu Westdeutschland bei den 17-Jährigen ein um 17 Prozent höherer Abiturientenanteil feststellen (17-Jährige West: 62,0 %; Ost: 78,8 %; siehe Abb. 5.3 und Abb. 5.5). Andererseits könnte die erhebliche Differenz der Abiturientenquote der räumlich Immobilen und der Quote aller Befragten (78,8 % - 59,8 % = 19 %) dieser Altersgruppe für Ostdeutschland darauf hinweisen, dass die Mehrheit der ausgefallenen Personen vor Ermittlung des Abiturs weggezogen ist.



Quelle: Mikrozensuspanel 1996-1999 (n = 132); eigene Berechnungen.

Abbildung 5.5: Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996, räumlich Immobile – Bundesländer mit zwölfjährigem Schulbesuch

Ausgehend von diesen Darstellungen des Vergleichs der Bildungsbeteiligung und Schulabschlüsse an den zeitlichen Rändern des Beobachtungsfensters 1996 und 1999 des Mikrozensuspanels werden im Folgenden die Bildungsverläufe dieser Altersgruppe differenzierter abgebildet. Wie aus Tabelle 5.2 für die Länder mit 13-jährigem Schulbesuch erkennbar, hängt die für 16-Jährige auffallend niedrige Abiturientenquote in hohem Maße mit dem Wechsel in berufliche Schulen zusammen.

Insgesamt erfolgen – auch für die anderen Altersgruppen – die meisten Schulwechsel von der gymnasialen Oberstufe zu beruflichen Schulen. Diese Schätzungen sind allerdings wegen der unberücksichtigten Ausfälle im Umfang von rund 17 Prozent zu niedrig. In Bezug auf die Ausfallproblematik zeigt Tabelle 5.2, dass unter den Ausfällen in Relation zu den Gesamtausfällen die Wegzüge aus weiter befragten Haushalten, d. h. Auszüge aus dem Elternhaus, überwiegen.

Der Versuch, am Beispiel der 17-Jährigen ohne Berücksichtigung der Ausfälle, einen „typischen“ Verlauf bis zum Abitur nachzuzeichnen, ergibt, dass bis zum Alter von 20 Jahren (1999) 50 Prozent die Schule mit dem Abitur verlassen haben. In welchem Umfang später noch Abschlüsse realisiert werden, die aber wegen des vierjährigen Beobachtungszeitraums nicht erfassbar sind, deuten die Verteilungen für die nächst älteren Altersgruppen an.

Lässt man die Ausfälle der Einfachheit halber einmal außer Acht und berechnet eine jahresbezogene Abiturquote auf Basis der um die Ausfälle reduzierten Zahl der Schulbesucher zum jeweiligen Erhebungszeitpunkt, zeigen sich anhand der Maximalwerte zwischen 61 und 69 Prozent plausible Werte, insofern die meisten Absolventen das Abitur im Alter von 19 bis 20 Jahren erreichen.

Tabelle 5.2: Bildungsweg der Besucher der gymnasialen Oberstufe in Bundesländern mit 13-jährigem Schulbesuch insgesamt und für die 1996 16- bis 19-Jährigen (Spaltenprozentwerte)

Bildungsverlauf	Insgesamt	darunter im Alter von ... Jahren (1996)			
		16	17	18	19
Gymnasiale Oberstufe 1996 (n = 100 %)	1.378	204	337	365	234
Ausfall 1996/97	9,1	7,4	4,7	8,2	13,7
<i>darunter: Auszug</i> <sup>1</sup>	4,8	2,0	2,4	4,4	9,0
Schulwechsel <sup>2</sup> 1996/97	19,2	22,1	15,1	13,4	13,7
<i>darunter: Berufliche Schule</i>	7,8	12,3	7,7	7,1	6,4
Abitur 1997	18,9	1,0	3,3	14,0	53,0
<i>Abiturquote 1996/97</i> <sup>3</sup>	20,8	1,1	3,4	15,2	61,4
Gymnasiale Oberstufe 1997	52,8	69,6	76,9	64,4	19,7
Ausfall 1997/98	4,6	2,9	5,3	7,4	2,1
<i>darunter: Auszug</i>	2,7	1,0	2,7	4,9	1,7
Schulwechsel 1997/98	5,3	7,4	4,2	8,5	3,0
<i>darunter: Berufliche Schule</i>	2,4	5,4	2,1	1,9	1,7
Abitur 1998	14,8	2,5	9,8	35,6	11,1
<i>Abiturquote 1997/98</i>	30,7	3,7	13,7	62,5	63,4
Zugänge <sup>4</sup>	3,2	3,4	2,4	1,1	
Gymnasiale Oberstufe 1998	31,3	60,3	59,9	14,0	3,4
Ausfall 1998/99	3,3	5,9	6,5	1,6	0,9
<i>darunter: Auszug</i>	2,2	2,9	4,7	1,4	0,4
Schulwechsel 1998/99	4,1	6,9	6,8	2,7	1,3
<i>darunter: Berufliche Schule</i>	1,5	2,5	3,3	0,3	0,4
Abitur 1999	14,1	16,7	37,1	7,7	1,3
<i>Abiturquote 1998/99</i>	50,4	30,6	69,4	62,2	50,0
Zugänge	6,9	2,9	0,6	0,3	
Gymnasiale Oberstufe 1999	10,8	33,8	10,1	2,2	0
Ausfälle (vor Abitur) insgesamt (%)	17,1	16,2	16,6	17,3	16,7
Abitur insgesamt (%)	47,8	20,1	50,1	57,3	65,4
Schulwechsel insgesamt (%)	28,5	36,3	26,1	24,7	17,9

<sup>1)</sup> „Auszüge“ umfassen hier nur Wegzüge aus weiter im Mikrozensuspanel befragten Haushalten. Die Differenz zum Insgesamt der Ausfälle beinhaltet auch Wegzüge gesamter Haushalte.

<sup>2)</sup> „Schulwechsel“ beinhalten alle Abgänge aus der gymnasialen Oberstufe, einschließlich kein Schulbesuch.

<sup>3)</sup> Die Abiturquote bezieht sich auf den Anteil der Schulbesucher zum jeweiligen Erhebungszeitpunkt verringert um den Anteil der Ausfälle; z. B. für 1996/97: 20,8 % = 18,9 %/(100 % - 9,1 %).

<sup>4)</sup> „Zugänge“ sind Besucher der gymnasialen Oberstufe, die nach 1996 diese Schulform verlassen haben (erfasst als „Schulwechsel“) und später wieder die gymnasiale Oberstufe besuchen.

*Quelle:* Mikrozensuspanel 1996-1999; eigene Berechnungen (ungewichtete Fallzahlen).

Tabelle 5.3 enthält die Verteilungen für die neuen Bundesländer mit zwölfjährigem Schulbesuch. Kumuliert man die Abituranteile, ergibt sich mit Ausnahme der 1996 16-Jährigen, dass die Mehr-

heit der Besucher der gymnasialen Oberstufe das Abitur bis zum Alter von 19 Jahren erwirbt. Der im Vergleich zu den anderen Ländern frühere Abschluss spiegelt sich auch in den etwas höheren jahresbezogenen Abiturquoten wider. In Bezug auf Ausfälle vor dem Abitur und Wechsel in andere Schulen (u. a. berufliche Schulen) unterscheiden sich die Verlaufsmuster in West und Ost nicht wesentlich. Aufgrund der geringen Fallzahlen in Tabelle 5.3 sind diese Vergleiche allerdings nur eingeschränkt möglich.

Das bei der Feststellung eines mit dem Abitur abgeschlossenen Besuchs der gymnasialen Oberstufe verwendete Verfahren setzt gültige Angaben zum allgemein bildenden Schulabschluss voraus. Dies trifft, wie oben im Fall der Besucher von Fachhochschulen und Hochschulen zu sehen, die erst zu einem späteren Zeitpunkt den Abschluss angegeben haben, nicht immer zu.

*Tabelle 5.3:* Bildungsverlauf der Besucher der gymnasialen Oberstufe in Bundesländern mit zwölfjährigem Schulbesuch insgesamt und für die 1996 16- bis 19-Jährigen (Spaltenprozentwerte)

Bildungsverlauf	Insgesamt	darunter im Alter von ... Jahren (1996)			
		16	17	18	19
Gymnasiale Oberstufe 1996 (n = 100 %)	231	44	87	67	11
Ausfall 1996/97	9,5	6,8	8,0	13,4	9,1
Schulwechsel 1996/97	16,9	15,9	12,6	10,4	18,2
<i>darunter: Berufliche Schule</i>	8,2	11,4	9,2	7,5	9,1
Abitur 1997	22,5		5,7	55,2	72,7
<i>Abiturquote 1996/97</i>	24,9		6,3	63,8	80,0
Gymnasiale Oberstufe 1997	51,1	77,3	73,6	20,9	0
Ausfall 1997/98	5,2	6,8	6,9	4,5	
Schulwechsel 1997/98	5,2	15,9	4,6	1,5	
Abitur 1998	21,2	4,5	43,7	13,4	
<i>Abiturquote 1997/98</i>	46,2	6,5	65,5	81,8	
Zugänge	3,0	4,5			
Gymnasiale Oberstufe 1998	22,5	54,5	18,4	1,5	
Ausfall 1998/99	3,9	6,8	4,6		
Schulwechsel 1998/99	1,7	4,5	2,3		
Abitur 1999	10,4	27,3	10,3	1,5	
<i>Abiturquote 1998/99</i>	55,8	57,1	75,0	100	
Zugänge	6,1	50,0			
Gymnasiale Oberstufe 1999	7,8	18,2	1,1	0	
Ausfälle (vor Abitur) insgesamt (%)	18,6	20,5	19,5	17,9	9,1
Abitur insgesamt (%)	54,1	31,8	59,7	70,1	72,7
Schulwechsel insgesamt (%)	23,8	36,4	19,5	11,9	18,2

*Quelle:* Mikrozensuspanel 1996-1999; siehe Anmerkungen zu Tab. 5.2.

Während oft angenommen wird, dass sich Klassifikationsfehler bei Querschnittsanalysen insgesamt aufheben, sind diese Probleme in Längsschnittuntersuchungen von zentraler Bedeutung. Wechselnde, d. h. zeitlich inkonsistente Angaben können zur Unter- oder Überschätzung der Häufigkeit von Statusübergängen führen. Treten diese verstärkt auf, ist damit zu rechnen, dass

dieses „Rauschen“ als substanzielle Veränderung interpretiert wird (Duncan 2000; Skinner 2000). Zur Beurteilung der Problematik selektiver Ausfälle im Mikrozensuspanel ist es daher sinnvoll, zunächst Fragen der Datenqualität der Bildungsangaben im Mikrozensus zu klären.

### **5.3 Zur Datenqualität der Bildungsangaben im Mikrozensus im Vergleich zur Bildungsstatistik**

Datenqualität bezieht sich allgemein darauf, in welcher Weise statistische Ergebnisse den mit der Datenerhebung verfolgten Zweck erfüllen. Zu den wichtigsten Aspekten zählen Schätzgenauigkeit und Reliabilität sowie Validität, aber auch die Verfügbarkeit und der Datenzugang (Körner und Schmidt 2006; Office of Management and Budget 2001: 1-2). Häufig steht bei Fragen der Datenqualität die Genauigkeit bzw. der Stichprobenfehler im Vordergrund. Jedoch dürfen bei der Frage der Datenqualität systematische Fehler keinesfalls ausgeblendet werden, da sie in vielen Umfragen größer als der Stichprobenfehler sein können (Särndal et al. 1997: 539).

Während zu den erwerbsstatistischen Themen des Mikrozensus solche Qualitätsuntersuchungen ansatzweise vorliegen (Dräther et al. 2001; Pöschl 1992; Rudolph 1998; Schupp et al. 1999; Statistisches Bundesamt 2006b), sind zu den bildungsstatistischen Angaben des Mikrozensus bisher kaum Einschätzungen möglich. Zu den wenigen Ausnahmen zählen insbesondere Untersuchungen zu den Auswirkungen der Freiwilligkeit der Auskünfte auf die Qualität der Bildungsangaben (Esser et al. 1989; Riede und Emmerling 1994), in Bezug auf den Hochschulabschluss frühere Vergleiche des Mikrozensus mit der Hochschulstatistik (Esser et al. 1989: 131) sowie zur beruflichen Weiterbildung ein Vergleich mit dem Berichtssystem Weiterbildung (Lois 2005).

#### **5.3.1 Methode und Abweichungsgründe**

Die Aufdeckung systematischer Fehler ist nicht einfach und setzt oftmals externe Datenquellen voraus. Im Idealfall werden verschiedene Verfahren kombiniert. Dazu gehören Pretests zur Kontrolle des Fragebogens und Wiederholungsbefragungen mit anderen Frageversionen bzw. differenzierten Nachfragen zu einzelnen Sachverhalten, die zugleich der Interviewerkontrolle dienen können. In wenigen Fällen sind Verknüpfungen von Umfragedaten mit Registerdaten und ein Abgleich der in beiden Datenquellen erfassten Merkmale möglich. Häufiger werden Randverteilungen mit anderen Daten verglichen. Bei den standardgemäß durchgeführten Analysen US-amerikanischer statistischer Ämter zur Datenqualität kommen diese Methoden häufig zum Einsatz (siehe unter anderem: Black et al. 2003; Brick et al. 1996; McGovern und Bushery 1999; NCES 1997; U.S. Census Bureau 2004).

Im Folgenden werden lediglich Randverteilungsvergleiche, also einfache deskriptive Vergleiche hochgerechneter Mikrozensusergebnisse mit der Bildungsstatistik durchgeführt. Dabei wird nicht vorausgesetzt, dass die Bildungsstatistik „wahre“ Werte liefert. Allerdings wird angenommen, dass die Ergebnisse dieser Totalerhebung deutlich weniger fehlerbehaftet sind als die des Mikrozensus. Diese Annahme erscheint gerechtfertigt.

Als Quellen systematischer Fehler kommen unklare Fragen oder Filterführungen im Fragebogen, Probleme der Erhebungs- bzw. Befragungsmethode, Interviewereinflüsse oder Verständnisschwierigkeiten und fehlerhafte Auskünfte sowie Antwortausfälle vonseiten der Befragten in Frage. Aus dem Vergleich mit der Bildungsstatistik kann nur die Gesamtabweichung („Total Survey Error“; siehe Biemer und Lyberg 2003) ermittelt werden. Inwiefern die Gesamtabweichungen mit dem Stichprobenfehler, unterschiedlichen Berichtszeiträumen, Unit- und Item-Nonresponse, verschiedenen Antwortfehlern (u. a. aufgrund sozialer Erwünschtheit) oder unterschied-

lichen Befragungsarten (Mode-Effekte) zusammenhängen, kann nicht geklärt werden. Zur ansatzweisen Kontrolle werden, sofern möglich, weitere Differenzierungen vorgenommen.

Die den Gesamtabweichungen zwischen Mikrozensus und Bildungsstatistik zugrunde liegenden Ursachen können wie folgt zusammengefasst werden:

- *Berichtszeitraum:* Während die Bildungsstatistik die Schülerzahlen zum Schulanfang eines Schuljahres (Herbst) und die Zahl der Absolventen zum Schuljahresende (Frühsommer) berichtet, beziehen sich die Mikrozensusergebnisse bis 2004 auf die Berichtswoche des (Folge-) Jahres, die i. d. R. in der letzten feiertagsfreien Aprilwoche liegt. Die Unterschiede, die auf die Differenz zwischen Herbst des Vorjahres und April zurückzuführen sind, dürften jedoch gering sein, da Schulartwechsel und Ausbildungsabbrüche hauptsächlich zum Schuljahresende stattfinden. Die Unterschiede der Absolventenzahlen in der Bildungsstatistik (Erhebungszeitpunkt z. B. Schuljahresende 1995/96) und im Mikrozensuspanel (hier z. B. April 1997) könnten dagegen überwiegend mit dem Auszug nach bestandem Abitur zusammenhängen.
- Zu berücksichtigen ist aber, dass im Mikrozensus nach dem „gegenwärtigen“ Schulbesuch gefragt wird. Bei einem bis in den Sommer reichenden Befragungszeitraum kann die Fokussierung auf den Interviewzeitpunkt dazu führen, dass Schüler, die in den Schulferien, d. h. nach Abschluss des Schuljahres befragt werden, Angaben zum nächsten Schuljahr (Herbst) statt zum vergangenen Schuljahr machen. Gegebenenfalls wird auch der zwischenzeitlich erreichte Schulabschluss statt der noch im April besuchten Schule angegeben. Grundsätzlich nicht auszuschließen ist auch der Fall, dass Angaben exakt zum Befragungszeitpunkt erfolgen, d. h. eventuell in den Schulferien kein Schulbesuch berichtet wird. Allerdings ist damit kaum zu rechnen, sofern von den befragten Haushalten auch in den Schulferien der Hauptstatus eines Schülers wahrgenommen wird.
- *Merkmalsunterschiede:* Die im Mikrozensus verwendeten Definitionen stimmen weitestgehend mit der Schulstatistik überein bzw. die jeweiligen Schulformen können vergleichbar abgegrenzt werden. Für den Vergleich des Mikrozensus mit der Schulstatistik ist dies wichtig, da somit konzeptionelle Abweichungen ausgeschlossen werden können. Drittvariablen liegen in der Bildungsstatistik nur sehr begrenzt vor. In dem Zusammenhang, dass in der Bildungsstatistik das Alter bzw. das Geburtsjahr nicht jedes Jahr direkt erhoben, sondern entsprechende Altersverteilungen teilweise geschätzt werden, kann es zu Abweichungen zum Mikrozensus kommen. Diese dürften jedoch eher marginal sein.
- *Populationsschätzung (Hochrechnung):* Frühere Vergleiche zwischen Mikrozensus und Volkszählung haben gezeigt, dass die laufende Bevölkerungsfortschreibung, an die die Mikrozensusdaten angepasst werden, den Bevölkerungsbestand mit zunehmendem Abstand zur Volkszählung überschätzt (siehe Rendtel und Schimpl-Neimanns 2001; Statistische Ämter des Bundes und der Länder 2004). Durch die Anpassung können somit Fehler in den Mikrozensus übertragen werden. Die Abweichungen zu den nicht extra angepassten Ergebnissen der Bildungsstatistik sollten sich bei Kontrolle weiterer Variablen, die wie beispielsweise das Alter nicht in der Anpassung berücksichtigt werden, jedoch in Form einer konstanten Verschiebung zeigen.
- *Definitionen im Mikrozensus:* Wie oben angesprochen, zählen im Mikrozensus ab 1991 ausschließlich Schüler allgemein bildender Schulen der Klassenstufen 11-13 zur gymnasialen Oberstufe. Schüler beruflicher Schulen mit den Bildungszielen Fachhochschulreife oder allgemeine Hochschulreife sind dagegen als Schüler beruflicher Schulen einzuordnen. Es ist davon auszugehen, dass diese Definition nicht dem Alltagsverständnis entspricht und deshalb in der Befragungspraxis nicht adäquat umgesetzt werden kann.



- *Befragungsart (Mode-Effekte)*: Bei den Besuchern der gymnasialen Oberstufe liegt der Anteil der schriftlichen Auskünfte bei 15 Prozent. Da die genauen Definitionen nur im Interviewerhandbuch zu finden sind, nicht aber im Selbstausfüller-Fragebogen, ist damit zu rechnen, dass schriftlich Befragte bei der Beantwortung der Frage nach dem Schulbesuch auf ihre Alltagsdefinition zurückgreifen. Geht man aber davon aus, dass die Definitionen selbst im mündlichen Interview nicht immer umfassend berücksichtigt werden, dürften die unterschiedlichen Befragungsarten nur geringe Effekte besitzen.
- Bei Oberstufenschülern liegen im Mikrozensus 1999 mit 82 Prozent überdurchschnittlich häufig Fremdauskünfte vor. Bei Proxy-Interviews wird häufig eine eingeschränkte Datenqualität angenommen. Ob dies auch in Bezug auf die Angaben zum Schulbesuch zutrifft, ist fraglich, denn anders als bei Proxy-Angaben zum Einkommen oder zu Arbeitsstunden (Dawe und Knight 1997) kann man in diesem Fall davon ausgehen, dass Eltern über die von ihren Kindern besuchte Schule gut informiert sind.

### 5.3.2 Absolventen mit allgemeiner Hochschulreife und Fachhochschulreife in der Bildungsstatistik und im Mikrozensuspanel

In den obigen Analysen wurde sowohl in Bezug auf die einzelnen Erhebungszeitpunkte bzw. Übergänge (vgl. Tab. 5.1) als auch auf altersspezifische „Abschlussquoten“ (vgl. Tab. 5.2 und 5.3) ein niedriger Anteil an Absolventen der gymnasialen Oberstufe festgestellt. Man kann annehmen, dass diese Untererfassung teilweise mit dem Wegzug von Abiturienten nach ihrem Schulabschluss zusammenhängt. Die Daten der amtlichen Bildungsstatistik erlauben zwar keine Längsschnittanalysen von Bildungsverläufen, sodass direkte Vergleiche mit Ergebnissen des Mikrozensuspanels ausgeschlossen sind. Jedoch liefert die amtliche Bildungsstatistik für bestimmte Zeitpunkte die für den Vergleich mit Ergebnissen des Mikrozensus benötigten Populationswerte, d. h. inklusive der räumlich mobilen Schüler bzw. Absolventen. Aus den Abweichungen der Ergebnisse des Mikrozensuspanels von den Ergebnissen der amtlichen Bildungsstatistik kann man auf Effekte der Stichprobenausfälle infolge räumlicher Mobilität schließen.

In Tabelle 5.4 werden für den Zeitraum 1996–1998 die hochgerechneten Zahlen der Absolventen der gymnasialen Oberstufe ohne Ausfallkorrektur den Ergebnissen der amtlichen Bildungsstatistik gegenübergestellt.<sup>36</sup> Obwohl im Mikrozensuspanel die an beruflichen Schulen erreichten Abschlüsse (Allgemeine Hochschulreife und Fachhochschulreife) in hohem Maße inkonsistente Ergebnisse zeigen (Schimpl-Neimanns 2005: 18ff.), werden sie dennoch berichtet, da Probleme der Abgrenzung zwischen gymnasialer Oberstufe und beruflichen Schulen vermutet werden. Unterscheidet man zunächst nicht nach der Schulart, unterschätzt das Mikrozensuspanel die Zahl der Absolventen mit Fachhochschulreife oder Hochschulreife um elf (1998) bis 14 Prozent (1997). Diese Abweichung entspricht näherungsweise dem Anteil räumlich mobiler Personen. Die naheliegende Vermutung, dass hierfür Ausfälle verantwortlich sind, muss allerdings noch geprüft werden. Während aber die Absolventen der gymnasialen Oberstufe (Fachhochschul-/Hochschulreife zusammen) insgesamt nur zwischen 0,7 (1998) bis 5,9 (1995) Prozent unterschätzt

36 Diese provisorische Hochrechnung verwendet die Gewichte der Querschnittsdaten. Dabei werden neben den Anpassungen der jährlichen Mikrozensus-Querschnittsergebnisse an die laufende Bevölkerungsfortschreibung, die sogenannte gebundene Hochrechnung (siehe Rendtel und Schimpl-Neimanns 2001), die Ziehungswahrscheinlichkeiten und die Quoten erfolgreich zusammengeführter Privathaushalte des Mikrozensuspanels berücksichtigt. (Für weitere Vergleiche mit anderen Gewichtungsfaktoren und Ausfallkorrekturen siehe Abbildung 5.10.)



werden, sind die Differenzen bei den beruflichen Schulen beträchtlich größer (1996: - 26 %; 1998: - 35 %).

Tabelle 5.4: Absolventen allgemein bildender und beruflicher Schulen mit Fachhochschulreife und Hochschulreife 1996 bis 1998 in der Bildungsstatistik und im Mikrozensuspanel (in 1.000)

Jahr	Schulwesen	Abschlussart	Mikro- zensuspanel	Bildungs- statistik	Relative Differenz (in %)
1996	Allgemein bildende Schulen	Fachhochschulreife	19,3	6,5	196,9
		Hochschulreife <sup>1</sup>	189,8	215,8	- 12,0
		Insgesamt	209,1	222,3	- 5,9
	Berufliche Schulen	Fachhochschulreife	35,0	66,8	- 47,6
		Hochschulreife	33,5	26,2	27,9
		Insgesamt	68,5	93,0	- 26,3
	Zusammen	Fachhochschulreife	54,2	73,3	- 26,1
		Hochschulreife	223,3	242,0	- 7,7
		Insgesamt	277,6	315,3	- 12,0
1997	Allgemein bildende Schulen	Fachhochschulreife	24,1	6,6	265,2
		Hochschulreife	189,7	219,6	- 13,6
		Insgesamt	213,8	226,2	- 5,5
	Berufliche Schulen	Fachhochschulreife	33,7	70,0	- 51,9
		Hochschulreife	31,6	27,3	15,8
		Insgesamt	65,3	97,2	- 32,8
	Zusammen	Fachhochschulreife	57,8	76,6	- 24,5
		Hochschulreife	221,3	246,9	- 10,4
		Insgesamt	279,1	323,4	- 13,7
1998	Allgemein bildende Schulen	Fachhochschulreife	27,9	8,2	240,2
		Hochschulreife	196,0	217,2	- 9,8
		Insgesamt	223,9	225,5	- 0,7
	Berufliche Schulen	Fachhochschulreife	34,7	73,9	- 53,0
		Hochschulreife	31,3	27,7	13,0
		Insgesamt	66,0	101,6	- 35,0
	Zusammen	Fachhochschulreife	62,6	82,2	- 23,8
		Hochschulreife	227,3	244,9	- 7,2
		Insgesamt	289,9	327,1	- 11,4

<sup>1)</sup> Unter „Hochschulreife“ sind Absolventen mit Allgemeiner und Fachgebundener Hochschulreife zusammengefasst.

*Quellen:* Bildungsstatistik: Statistisches Bundesamt 1999b: 108. (Jahresangaben = Entlassungsjahr.) Mikrozensuspanel 1996-1999: Absolventen/Besucher der gymnasialen Oberstufe mit festgestelltem Abschluss; gebundene Hochrechnung ohne Ausfallkorrektur; Bevölkerung in Privathaushalten am Hauptwohnsitz; eigene Berechnungen.

Markante Unterschiede treten zutage, wenn man sowohl nach Abschlussart als auch nach Schulart differenziert. Die Abschlüsse mit Fachhochschulreife werden bei den Absolventen der gymnasialen Oberstufe im Mikrozensuspanel mit über 200 Prozent gravierend überschätzt, bei den beruflichen Schulen dagegen um rund 50 Prozent unterschätzt. Umgekehrt werden die Absolventenzahlen der an allgemein bildenden Schulen erworbenen Hochschulreife zwischen zehn Prozent (1998) und 14 Prozent (1997) unterschätzt, während bei beruflichen Schulen eine Überschätzung zwischen 13 Prozent (1998) und 28 Prozent (1996) vorliegt.

Die den Abschlüssen mit Fachhochschulreife und den Absolventen beruflicher Schulen zugrunde liegenden Fallzahlen des Mikrozensuspanels sind zwar recht klein, doch selbst unter Berücksichtigung der dadurch eingeschränkten Aussagekraft weisen diese Vergleiche und insbesondere die starke Übererfassung des Fachhochschulreifeabschlusses an der gymnasialen Oberstufe darauf hin, dass im Mikrozensus vermutlich eine Reihe von Besuchern beruflicher Schulen, an denen die Fachhochschulreife oder Hochschulreife erworben werden kann, bei der Befragung angeben, die gymnasiale Oberstufe zu besuchen.

Die Unterschätzung der Zahl der Absolventen mit Hochschulreife aus der gymnasialen Oberstufe um zehn bis 14 Prozent korrespondiert näherungsweise mit den Anteilen der Ausfälle 18- bzw. 19-jähriger Besucher der gymnasialen Oberstufe 1996 (vgl. Tab. 5.2 und 5.3). Der Zusammenhang zwischen der Unterschätzung der Abschlüsse und den Ausfällen wird aber weiter unten untersucht werden.

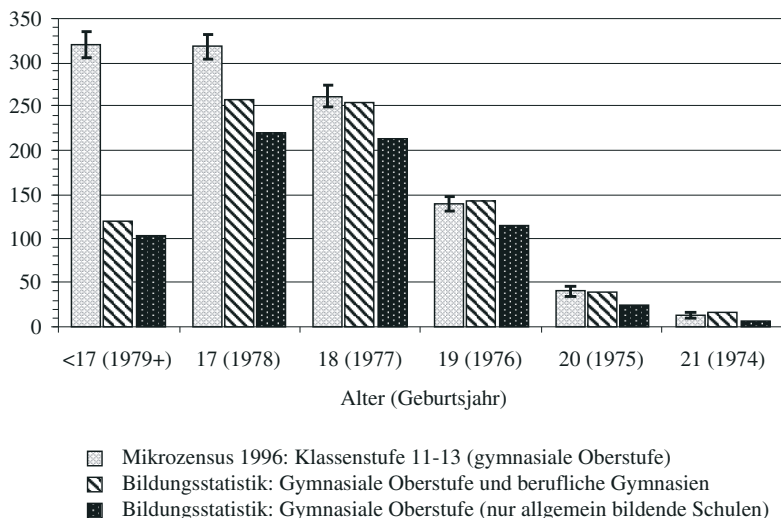
### **5.3.3 Vergleiche zum Besuch der gymnasialen Oberstufe im Mikrozensus und in der Bildungsstatistik**

Als Hauptgrund für die Abweichungen zwischen Mikrozensus und Bildungsstatistik in Bezug auf Schüler der gymnasialen Oberstufe wird angenommen, dass diese auf Begriffe und Definitionen zurückzuführen sind, die sich in der Befragungspraxis nicht unmittelbar umsetzen lassen. Zur Überprüfung dieser Vermutung stehen daher Vergleiche des Mikrozensus 1996 mit der Bildungsstatistik im Zentrum dieses Abschnitts. Sie werden durch Vergleiche früherer Mikrozensusdaten ergänzt.

Gemäß der Bildungsstatistik besuchten im Schuljahr 1995/96 679.900 Schüler im Alter bis zu 21 Jahren die gymnasiale Oberstufe (siehe Abb. 5.6). Der Mikrozensus 1996 weist dagegen hochgerechnet 1.093.600 Schüler aus. Die Differenz von über 413.000 Schülern entspricht einer Übererfassung von insgesamt rund 60 Prozent. Bei den Gesamtwerten des Mikrozensus sind die 95%-Konfidenzintervalle ausgewiesen.<sup>37</sup> Die Differenzen zur Bildungsstatistik sind offensichtlich nicht durch den Stichprobenfehler zu erklären. Die Überschätzung reduziert sich insgesamt auf ungefähr die Hälfte (32 %), wenn zu den Schülern der Sekundarstufe II allgemein bildender Schulen der Bildungsstatistik die Schüler beruflicher Gymnasien hinzugerechnet werden. In diesem Fall schließt das Konfidenzintervall des Mikrozensus bei den 18- bis 21-Jährigen die entsprechenden Gesamtwerte der Bildungsstatistik ein. Dies deutet darauf hin, dass im Mikrozensus Besucher beruflicher Gymnasien entgegen den Definitionen des Mikrozensus als Schüler der gymnasialen Oberstufe erfasst sind. Für diese Annahme spricht auch, dass durch die Herausnahme der Schüler beruflicher Gymnasien (Fachoberschulen, Fachgymnasien und Berufs-/Technische Oberschulen) aus den Gesamtwerten der Berufsschulstatistik die Anpassung des Mikrozensus an

37 Zur Berechnung der Stichprobenfehler bei der hier verwendeten gebundenen Hochrechnung (Anpassung an die Bevölkerungsfortschreibung) siehe Rendtel und Schimpl-Neimanns (2001: 103f.). Dieses Verfahren ist nur mit den Scientific Use Files des Mikrozensus ab 1996 anwendbar.

die Berufsschulstatistik verbessert werden kann (hier ohne Nachweis; siehe Schimpl-Neimanns 2005: 28-29).



#### Quellen:

Bildungsstatistik: Gymnasiale Oberstufe allgemein bildender Schulen: Statistisches Bundesamt 1997a: 24, 36-37. Berufliche Schulen: Statistisches Bundesamt 1997b: 17. (Gymnasiale Oberstufe: N = 679.900; gymnasiale Oberstufe inkl. berufliche Gymnasien: N = 830.300).

Mikrozensus 1996 (faktisch anonymisierte 70%-Substichprobe); Bevölkerung am Ort der Hauptwohnung; hochgerechnete, an die Bevölkerungsfortschreibung angepasste Fallzahlen (N = 1.093.600; ungewichtet: n = 6.753); 95%-Konfidenzintervall: I; eigene Berechnungen.

Abbildung 5.6: Besucher der gymnasialen Oberstufe im Mikrozensus 1996 und in der Bildungsstatistik (Schuljahr 1995/96) (in 1.000)

Bei den unter 18-Jährigen, insbesondere bei den unter 17-Jährigen bestehen dennoch weiterhin erhebliche Übererfassungen, die offensichtlich nicht mit der Fehlklassifikation beruflicher Gymnasien zu erklären sind. Diese Abweichungen könnten auf das oben genannte Problem des Berichtszeitraums bzw. auf die Fokussierung auf den Interviewzeitpunkt und Antworten zum Schulbesuch des neuen Schuljahres zurückzuführen sein. Sie könnten aber auch damit zusammenhängen, dass es den Befragten schwer fällt, zwischen den Sekundarstufen I und II (Klassenstufen 5-10 vs. 11-13) zu unterscheiden und die zur Klassenstufe 11-13 in Klammern genannte Ergänzung „gymnasiale Oberstufe“ allgemein als „Gymnasium“ missverstanden wird.<sup>38</sup> Da die Bildungsstatistik die Altersgruppen, bei denen Überschneidungen zwischen den Sekundarstufen I und II auftreten, nur zusammengefasst in nach oben bzw. unten offenen Flügelklassen ausweist, lassen sich diese Vermutungen nicht weiter verfolgen.

Die genannten Befunde zeigen sich auch für die Erhebungszeitpunkte 1997 bis 1999 (siehe Schimpl-Neimanns 2005: 56-58) und weisen somit auf ein systematisches Problem hin. Als Ur-

38 Im Mikrozensus 1996 enthält die Flügelklasse der unter 17-Jährigen auch den Geburtsjahrgang 1981. Dass hochgerechnet rund 18.400 Schüler, die im Herbst 1995 etwa 14 bis 15 Jahre alt waren, einen Besuch der gymnasialen Oberstufe angeben, könnte als Indiz für eine unscharfe Differenzierung zwischen den Sekundarstufen I und II gewertet werden.

sachen wurden oben auch andere potenzielle Zusammenhänge zwischen Befragungsart und Verteilungsabweichung angesprochen. Um diese versuchsweise zu kontrollieren, wurde eine logistische Regression zum Verhältnis des Besuchs der gymnasialen Oberstufe vs. einer beruflichen Schule mit soziodemografischen Variablen und den Befragungsarten Selbstausfüller und Proxy-Interview geschätzt. Mit den Daten des Mikrozensus 1999 ergaben sich keine statistisch signifikanten Interaktionen zwischen den Variablen der Befragungsart und den anderen erklärenden Variablen (hier ohne Darstellung). Damit kann zwar nicht ausgeschlossen werden, dass solche Effekte dennoch vorhanden sind, für eine Überprüfung von Mode-Effekten liegen jedoch im Mikrozensus keine geeigneten Informationen vor. Von entscheidender Bedeutung wäre die Kenntnis des Interviewzeitpunkts.

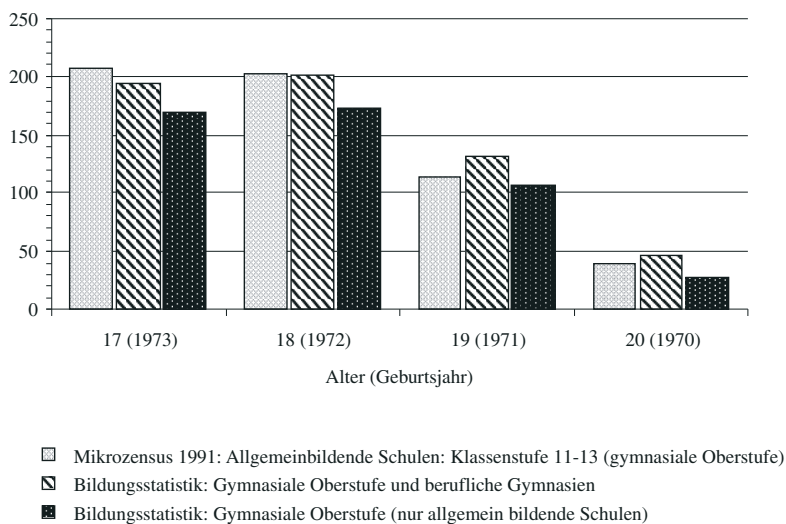
Um die Plausibilität der Erklärung der Verteilungsunterschiede trotz fehlender methodischer Informationen zu testen, kann auf weitere Mikrozensusdaten zurückgegriffen werden. Bis 1990 wurde der Schulbesuch nach Schularten erfragt. Alle allgemein bildenden und beruflichen Schulen mit dem Bildungsziel allgemeiner Hochschulreife (Abitur), fachgebundener Hochschulreife oder Fachhochschulreife wurden in der Kategorie „Gymnasium/Fachoberschule“ zusammengefasst. Trifft die Annahme der Fehlklassifikation der beruflichen Gymnasien zu, ist zu erwarten, dass die Abweichungen zwischen Bildungsstatistik und Mikrozensus 1989 geringer ausfallen. Zum Vergleich werden zunächst die Mikrozensusergebnisse des Jahres 1991 herangezogen. Da im Mikrozensus 1989 die Abgrenzung der Sekundarstufe II nur näherungsweise mittels des Alters möglich ist, wird im Folgenden auf die nach unten offene Flügelklasse sowie auf die Altersgruppe der 21-Jährigen, die für einige Schulformen in der Bildungsstatistik ebenfalls zusammengefasste Jahrgänge enthält, verzichtet. Außerdem beziehen sich die Gegenüberstellungen auf Westdeutschland, d. h. nur auf das frühere Bundesgebiet.

Der Vergleich des Mikrozensus 1991 mit der Bildungsstatistik (siehe Abb. 5.7) zeigt in der Tendenz geringere, aber im Muster ähnliche Abweichungen wie beim Mikrozensus 1996.<sup>39</sup> Insgesamt wird die Zahl der gymnasialen Oberstufenschüler um 18 Prozent (= 563.800/477.500) überschätzt. Zur Gesamtabweichung trägt hauptsächlich die Übererfassung von rund 22 Prozent bei den 17-Jährigen bei. Dies verweist wiederum auf die Probleme der Unterscheidung der Sekundarstufen bzw. die Fokussierung auf den Interviewzeitpunkt.

Wie beim Mikrozensus 1996 sind, mit Ausnahme der 19-Jährigen, die altersspezifischen Abweichungen zur Bildungsstatistik geringer, wenn zu den Besuchern der gymnasialen Oberstufe auch die Schüler beruflicher Gymnasien gezählt werden. Bei den über 18-Jährigen kehrt sich allerdings die Übererfassung in eine Untererfassung von rund 13 Prozent um. Insgesamt ist die Differenz zwischen Mikrozensus und Bildungsstatistik von 1,3 Prozent vernachlässigbar, wenn man in der Bildungsstatistik die Schüler beruflicher Gymnasien hinzunimmt.

---

39 Da die neuen Fragen zum Schulbesuch erstmals 1991 gestellt wurden, ist nicht auszuschließen, dass die Interviewer teilweise noch die alten Definitionen zum Schulbesuch angewendet haben. Dies könnte ein Grund für die 1991 im Vergleich zu 1996 geringeren Abweichungen zwischen Mikrozensus und Bildungsstatistik und das im Vergleich zum Mikrozensus 1989 ähnliche Abweichungsmuster sein.



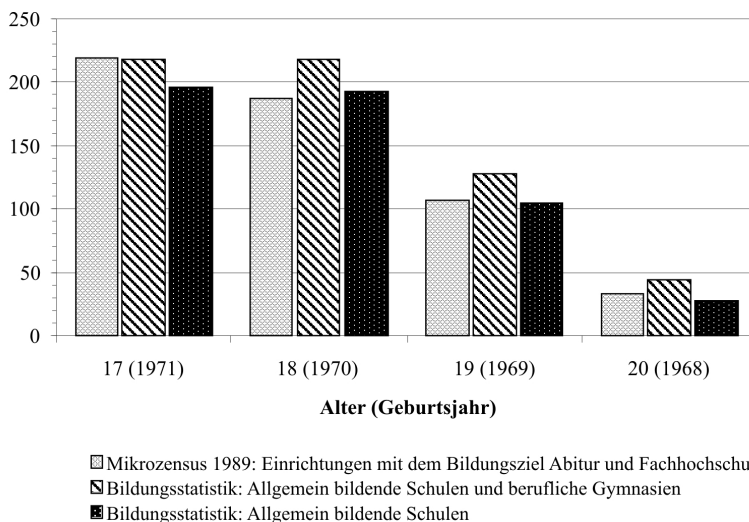
### Quellen:

Bildungsstatistik: Gymnasiale Oberstufe allgemein bildender Schulen (N = 477.500); gymnasiale Oberstufe inkl. berufliche Gymnasien (N = 571.500); Statistisches Bundesamt 1992.

Mikrozensus 1991 (faktisch anonymisierte 70 %-Substichprobe); Bevölkerung am Ort der Hauptwohnung; hochgerechnete, an die Bevölkerungsfortschreibung angepasste Fallzahlen (N = 563.800; ungewichtet n = 3.578); eigene Berechnungen.

Abbildung 5.7: Besucher der gymnasialen Oberstufe im Mikrozensus 1991 und in der Bildungsstatistik (Schuljahr 1990/91) – Westdeutschland (in 1.000)

Im Unterschied zu den Erhebungen ab 1991 sind 1989 berufliche Gymnasien den allgemein bildenden Schulen der Sekundarstufe II zugeordnet. Der Besuch der Gymnasialzüge an Gesamt- und Sonderschulen (Sekundarstufe II) war 1989 ebenfalls in der Kategorie „Gymnasium/Fachoberschule“ einzutragen. Die genauen Zuordnungen werden wiederum lediglich im Interviewerhandbuch beschrieben. Mit dem Mikrozensus 1989 werden 547.700 Schüler im Alter von 17 bis 20 Jahren geschätzt, die Schulen mit dem Bildungsziel des Abiturs oder der Fachhochschulreife besuchen (siehe Abb. 5.8). Der hierzu vergleichbare Gesamtwert der Bildungsstatistik beträgt 609.400 Schüler. Er umfasst die der gymnasialen Oberstufe zugeordneten allgemein bildenden Schulen sowie berufliche Gymnasien. Damit liegt im Mikrozensus 1989 insgesamt eine Untererfassung von zehn Prozent vor. Im Vergleich zur Übererfassung der gymnasialen Oberstufe im Mikrozensus 1991 (18 %) und insbesondere zur gravierenden Übererfassung dieser Gruppe im Mikrozensus 1996 (60 %) ist die (absolute) Abweichung des Mikrozensus 1989 vom Sollwert der Bildungsstatistik deutlich geringer. Dies unterstützt die obige Vermutung, dass die im Mikrozensus ab 1991 verwendete Abgrenzung der gymnasialen Oberstufe mit der Zuordnung beruflicher Gymnasien zu beruflichen Schulen in der Befragungspraxis schwer umzusetzen ist und eine wichtige Quelle der Verteilungsabweichungen zur Bildungsstatistik darstellt.



#### Quellen:

Bildungsstatistik: Allgemein bildende Schulen (N = 520.400); allgemein bildende Schulen inkl. berufliche Gymnasien (N = 609.400); Statistisches Bundesamt 1990.

Mikrozensus 1989 (faktisch anonymisierte 70 %-Substichprobe); Bevölkerung am Ort der Hauptwohnung; hochgerechnete, an die Bevölkerungsfortschreibung angepasste Fallzahlen (N = 547.700; ungewichtet n = 3.338); eigene Berechnungen.

**Abbildung 5.8:** Schüler allgemein bildender Schulen und beruflicher Gymnasien mit dem Bildungsziel Abitur und Fachhochschulreife im Mikrozensus 1989 und in der Bildungsstatistik (Schuljahr 1988/89) (in 1.000)

In Bezug auf die Frage, ob die Abweichungen zwischen dem Mikrozensus ab 1991 und der Bildungsstatistik bei den unter 18-Jährigen mit dem Problem der Fokussierung auf den Interviewzeitpunkt oder mit der schwierigen Unterscheidung der Sekundarstufen I und II (bzw. Klassenstufen 5-10 vs. 11-13) zusammenhängen, ist der Vergleich der Verteilungsunterschiede der Mikrozensus 1989 und 1991 zur Bildungsstatistik aufschlussreich. Verwendet man als Sollzahlen der Bildungsstatistik die Gesamtwerte der Schüler allgemein bildender Schulen der gymnasialen Oberstufe einschließlich beruflicher Gymnasien, zeigt sich für die 17-Jährigen folgender Befund: Während im Mikrozensus 1991 eine Übererfassung des Gesamtwerts der Bildungsstatistik um rund sieben Prozent vorliegt,<sup>40</sup> wird der Sollwert der Bildungsstatistik mit dem Mikrozensus 1989 nur um 0,5 Prozent überschätzt. Dies ist vernachlässigbar. Da in beiden Mikrozensus-erhebungen nach dem „gegenwärtigen Schulbesuch“ gefragt wird, wäre davon auszugehen, dass Effekte der Fokussierung auf den Interviewzeitpunkt gleichermaßen sowohl 1989 als auch 1991 auftreten. Wenn aber bei den 17-Jährigen 1989 im Unterschied zu 1991 keine Übererfassung festzustellen ist, kann dies 1989 nicht mit Fokussierungseffekten zusammenhängen. Vergleichbar zum Mikrozensus 1989 enthalten die Zahlen der Bildungsstatistik des Schuljahres 1988/89 neben den 17- bis 20-jährigen Schülern beruflicher Gymnasien die gleichaltrigen Schüler allgemein bildender Schulen, die der gymnasialen Oberstufe zuzurechnen sind. Zwar wird dabei die Sekundarstufe II nur näherungsweise durch die Altersgliederung abgegrenzt, dennoch spricht das Er-

40 Beim Mikrozensus 1996 (einschließlich neue Bundesländer) beträgt die Übererfassung der 17-jährigen Schüler der gymnasialen Oberstufe einschließlich beruflicher Gymnasien rund 23 Prozent.

gebnis dafür, dass die in den Mikrozensen ab 1991 bei den jüngeren, unter 18-jährigen Schülern der gymnasialen Oberstufe festgestellten starken Übererfassungen kaum auf Fokussierungseffekte, sondern auf eine nicht adäquate Unterscheidung der Sekundarstufen I und II zurückzuführen sind.

Die Untererfassung der 17- bis 20-jährigen Schüler allgemein bildender Schulen der gymnasialen Oberstufe und beruflicher Gymnasien im Mikrozensus 1989 um zehn Prozent ist jedoch nicht zu vernachlässigen. Bei der Frage, womit diese Untererfassung zusammenhängen könnte, ist an die Anmerkungen in Unterabschnitt 5.3.1 anzuknüpfen. Wie dort ausgeführt, werden die genauen Zuordnungen der Schulformen zu den Antwortkategorien im Mikrozensus 1989 nur im Interviewerhandbuch genannt. Sowohl im Interviewerbogen als auch im Selbstausfüllerbogen ist lediglich der Sammelbegriff „Gymnasium/Fachoberschule“ angegeben. Nicht explizit genannte Schulen können auf diese Weise leicht übersehen werden. Für die Untererfassung dürfte somit maßgeblich die unvollständige Antwortliste verantwortlich sein.

### 5.3.4 Antwortstabilität der Angaben zum allgemein bildenden Schulabschluss

Wie bisher zu sehen war, treten im Mikrozensus systematische Fehler in Form von Klassifikationsfehlern etc. auf. Die Panelangaben des Mikrozensus eröffnen neue Möglichkeiten zur Beurteilung der Antwortkonsistenz bzw. Antwortstabilität im Zeitverlauf. Dabei können geringe Antwortstabilitäten ergänzend zu den vorherigen Analysen auf Probleme bei der Erhebung von Bildungsabschlüssen hinweisen. Trifft die zum Schulbesuch gezogene Schlussfolgerung zu, dass Schüler der gymnasialen Oberstufe und beruflicher Schulen nicht den Definitionen des Mikrozensus entsprechend unterscheidbar sind, ist zu erwarten, dass sich dies auch bei den in diesen Schulformen erworbenen typischen Abschlüssen der allgemeinen bzw. fachgebundenen Hochschulreife (gymnasiale Oberstufe) vs. Fachhochschulreife (berufliche Schulen) widerspiegelt. Aus Mustern von Übergangswahrscheinlichkeiten lässt sich allgemein schließen, dass den Befragten bestimmte Antwortkategorien ähnlich erscheinen. Demnach sind beim Fachhochschulreifeabschluss im Vergleich zu anderen Abschlüssen eine geringere zeitliche Stabilität und eine höhere Übergangswahrscheinlichkeit zum Abitur zu erwarten.

Zur Qualität von Bildungsangaben kann nur auf sehr wenige Erkenntnisse zurückgegriffen werden (Reimer 2001: 114). In der Test-Retest Studie zum ALLBUS 1984 (n=154) wurden drei Interviews im Abstand von vier Wochen durchgeführt. Für den allgemeinen Schulabschluss wurde eine Stabilität von 89 Prozent, für die berufliche Qualifikation nur eine Stabilität von 72 Prozent ermittelt (Porst und Zeifang 1987: 196). Auf der Basis einer zweimaligen Befragung 89 verheirateter und geschiedener Ehepaare im Abstand von etwa einem halben Jahr, mit der sowohl Proxy-Effekte als auch Rückerinnerungseffekte untersucht wurden, berichtet Babka von Gos-towski (1995) für den Schulabschluss eine Gesamtstabilität von 72 Prozent. Orientiert man sich an diesen zwei Studien, kann eine Antwortstabilität von 70 Prozent als Minimum bezeichnet werden.

Mit dem Mikrozensuspanel liegt für diese Fragestellungen erstmals ein wesentlich umfangreicherer Datensatz vor (vgl. Herter-Eschweiler 2003: 243f.). Einschränkend muss jedoch bemerkt werden, dass keine Information über die Auskunft gebende Person verfügbar ist. Die Proxy-Angaben wurden zwar erstmals 1999 im Rahmen der EU-Arbeitskräftestichprobe erhoben, jedoch aus Datenschutzgründen nicht in das File aufgenommen. Im Folgenden werden nur räumlich immobile Personen im Alter von 25 bis 50 Jahren ohne Schulbesuch betrachtet, bei denen man von einer abgeschlossenen Schulkarriere ausgehen kann. Mit der oberen Altersgrenze werden der Einfachheit halber die besonderen Aspekte der Freistellung von der Auskunftspflicht für über



51-Jährige ausgeblendet. Ebenfalls vereinfacht die Konzentration auf räumlich Immobile bzw. der Ausschluss wegziehender Personen die Ergebnisdarstellung. Diese Selektion dürfte für die Frage konsistenter Antworten unerheblich sein. Für rund 29.000 Personen werden insgesamt rund 87.000 Übergänge beobachtet, d. h. aufgrund von zeitlich inkonsistenten Angaben liegen Mehrfachantworten vor.

Tabelle 5.5 zeigt für den Abschluss der allgemeinen bzw. fachgebundenen Hochschulreife, dem typischen Abschluss der gymnasialen Oberstufe, eine Antwortstabilität von 85 Prozent. Für Personen mit Fachhochschulreife, die zumeist an beruflichen Schulen erworben wird, wird lediglich eine Wahrscheinlichkeit von 50 Prozent festgestellt, dass dieser Abschluss auch bei der nächsten Befragung genannt wird. Die Übergangswahrscheinlichkeit zur Angabe Abitur beträgt 23 Prozent. Diese überdurchschnittlich inkonsistenten Abschlussangaben korrespondieren mit der oben bereits festgestellten Schwierigkeit von Befragten, den Besuch beruflicher und allgemein bildender Schulen der Sekundarstufe II zu differenzieren.

*Tabelle 5.5:* Antwortvariabilitäten zum allgemeinen Schulabschluss im Mikrozensuspanel 1996-1999 (Übergangswahrscheinlichkeiten zwischen den Zeitpunkten t und t+1) – Zeilenprozentwerte

Bildungsangabe (t) (m Beobachtungen; in Prozent)		Bildungsangabe (t+1)							
		0	1	2	3	4	5	8	9
0	Kein Abschluss (1,7 %)	<b>46,8</b>	40,0	2,0	2,8	0,3	1,6	1,1	5,5
1	Volks-/Hauptschule (40,3 %)	1,5	<b>87,4</b>	1,8	5,3	0,5	0,7	0,5	2,4
2	Polytechn. Oberschule (14,6 %)	0,2	5,9	<b>83,9</b>	5,1	1,5	1,5	0,2	1,7
3	Realschule (20,1 %)	0,2	10,0	3,1	<b>79,0</b>	3,0	2,4	0,4	1,9
4	Fachhochschulreife (4,2 %)	0,1	4,7	5,3	13,9	<b>50,5</b>	23,3	0,3	1,9
5	Hochschulreife (15,2 %)	0,2	1,6	1,3	3,2	6,5	<b>85,2</b>	0,4	1,7
8	Abschl. vorh., o. Ang. (0,5 %)	4,4	41,3	5,4	17,6	3,9	9,1	<b>7,2</b>	11,1
9	Keine Angabe (3,3 %)	2,8	36,6	8,3	14,4	3,0	9,8	2,6	<b>22,4</b>
Insgesamt (m = 87.234; 100 %)		1,6	40,9	14,4	20,4	4,3	15,1	0,5	2,8

*Quelle:* Faktisch anonymisiertes Mikrozensuspanel 1996-1999; räumlich immobile Personen im Alter von 25 bis 50 Jahren ohne Schulbesuch (ungewichtete Fallzahlen; n = 29.078 Personen).

Hohe Antwortkonsistenzen können des Weiteren für die Abschlüsse der Volks- bzw. Hauptschule, der Polytechnischen Oberschule und der Realschule festgestellt werden. Personen, die zunächst angegeben haben, keinen allgemein bildenden Abschluss zu besitzen bzw. ihren Abschluss nicht genannt haben, wechseln im Erhebungszeitraum 1996-1999 häufig zur Angabe eines Hauptschulabschlusses (siehe Kategorien 0, 8 und 9 in Tab. 5.5). Auch wenn dies wegen der zugrunde liegenden kleinen Fallzahlen nur eingeschränkt zu interpretieren ist, deutet es darauf hin, dass die Panelangaben zur Korrektur des Item-Nonresponse bzw. zur Überprüfung der Angabe „kein Abschluss“ genutzt werden können.



## 5.4 Analysen zum Ausfall der Besucher der gymnasialen Oberstufe bis zur Ermittlung des Abiturs

Neben den Stichprobenausfällen sind, das haben die obigen Vergleiche gezeigt, inkonsistente und unplausible Angaben im Mikrozensus eine nicht zu vernachlässigende Ursache der Verteilungsabweichungen zwischen Mikrozensus und Bildungsstatistik. Diese Probleme der Datenqualität müssen jedoch bei der nun folgenden Bearbeitung der zentralen Frage selektiver Ausfälle außer Acht bleiben.

Wie im dritten Kapitel dargestellt, werden bei der Analyse selektiver Ausfälle für die Formulierung der Teile des Struktur- und Ausfallmodells (siehe S. 27) des log-linearen Modells substanzwissenschaftliche Theorien benötigt. Beim Strukturmodell wird auf theoretische Ansätze zur Bildungsungleichheit und beim Ausfallmodell auf den migrationstheoretischen Ansatz sowie die Ergebnisse zum Auszug aus dem Elternhaus (siehe Kap. 4) Bezug genommen. Diese Ansätze werden im nächsten Abschnitt vorgestellt. Darauf aufbauend folgen statistische Modelle, zunächst ohne Berücksichtigung zeitabhängiger Ausfallmuster, dann für Verläufe.

### 5.4.1 Ansätze zur Erklärung sozial ungleicher Bildungsbeteiligung

In der Sozialstrukturforschung zählen die Theorie sozialer Reproduktion und die mikrosoziologische Theorie für die Wahl der Schulausbildung zum Kern der Erklärungsansätze. Sie setzen für die Erklärung sozial ungleicher Bildungsbeteiligung jeweils spezifische Schwerpunkte.

Die Theorie kultureller Reproduktion (Bourdieu und Passeron 1971; Bourdieu 1973, 1982) weist darauf hin, dass die Verteilung der Schüler auf verschiedene Schulformen keineswegs als klassenneutral betrachtet werden kann. Da sich die für den Schulerfolg relevanten Kriterien an der Elitekultur orientieren und Kinder der unteren Klassen nicht über entsprechende kulturelle Ressourcen („kulturelles Kapital“) verfügen, schneiden sie in der schulischen Selektion schlechter ab als die oberen Klassen. Darüber hinaus besitzen die unteren Klassen oft kein ausreichendes eigenes Wissen über die Bedeutung der verschiedenen Bildungsgänge und lassen sich deshalb bei der Bewertung der Bildungsziele stärker als die oberen Klassen von schulischen Leistungskriterien leiten. Die herkunftsspezifisch ausgeprägten Einstellungen und Erwartungshaltungen gegenüber der Bildung bestimmen auch die Wahl von Bildungsgängen und führen nach Bourdieu (1973: 106) letztlich zur „Selbsteliminierung“ der unteren Klassen.

Boudon (1974) greift die von Bourdieu betonte Bedeutung des kulturellen Kapitals für den Bildungserfolg auf, betrachtet dagegen aber die Bildungsbeteiligung als Ergebnis einer quasi-ökonomischen Investitionsentscheidung.<sup>41</sup> Insbesondere das in der Familie vorhandene kulturelle Kapital hängt eng mit dem Schulerfolg bzw. den Schulleistungen zusammen, die eine wichtige Zugangsvoraussetzung für weiterführende Schulen sind (primärer Herkunftseffekt). Die Bewertung der Schulbesuchsalternativen erfolgt nach ihren Nutzen und den Motiven – z. B. in Bezug auf die spätere Verwertbarkeit im Berufsleben oder hinsichtlich des Erreichens eines der Herkunftsfamilie entsprechenden Status – und den mit dem Schulbesuch verknüpften Kosten – z. B. Schulkosten oder entgangene Einkommen (sekundärer Herkunftseffekt).

Gemäß Boudon legt das Bildungssystem für die individuellen Entscheidungen über die Bildungskarriere den Handlungsrahmen idealtypisch in der Weise fest, dass es die Schüler bzw. de-

41 Für Anwendungen, kritische Diskussionen und Weiterentwicklungen siehe z. B.: Breen und Goldthorpe 1997; Erikson und Jonsson 1996; Esser (1999: 265-274) und Mare 1980, 1981. Auf die dabei verfolgten unterschiedlichen Schwerpunktsetzungen (siehe u. a. Kristen 1999) wird in dieser Arbeit nicht eingegangen, da eine empirische Überprüfung mit den hier verwendeten Daten nicht möglich ist.

ren Familien zu bestimmten Selektionszeitpunkten wiederholt vor Entscheidungen über das Fortsetzen eines Bildungsgangs stellt. Je häufiger solche Entscheidungen gefällt werden müssen, umso stärker wirken insgesamt die sekundären Herkunftseffekte.

Zudem ist anzunehmen, dass jüngere Schüler stärker unter dem Einfluss der Herkunftsfamilie stehen als ältere Schüler (Blossfeld und Shavit 1993). Für zeitlich nachgelagerte Selektionsstufen wie die der gymnasialen Oberstufe ist deshalb und aufgrund der vorherigen Bildungsentscheidungen mit einer geringeren bzw. abnehmenden sozialen Ungleichheit zu rechnen. Dies zeigen auch verschiedene Untersuchungen zur Bildungsungleichheit in Deutschland (Blossfeld und Shavit 1993; Henz und Maas 1995; Hillmert und Jacob 2005; Müller und Haun 1994).

Einerseits ist bei der Untersuchung des Verlaufs vom Besuch der gymnasialen Oberstufe bis zum Abschluss mit dem Abitur davon auszugehen, dass es sich, etwa im Unterschied zu Real- und Hauptschülern, um eine vergleichsweise leistungshomogene Schülerpopulation handelt, insofern diese Gruppe die ersten zehn Bildungsjahre erfolgreich durchlaufen hat. Andererseits umfasst die Population auch Absolventen der Realschulen, welche am Ende der Sekundarstufe I die Möglichkeit zum Wechsel in die Sekundarstufe II bzw. in die gymnasiale Oberstufe wahrgenommen haben.

Für diese Gruppe der Schulformwechsler ist mit unterdurchschnittlichen Erfolgswahrscheinlichkeiten für das Erreichen des Abiturs zu rechnen (Pötzsch 1982). Im Mikrozensuspanel ist diese Gruppe zwar mangels Informationen über die vor der gymnasialen Oberstufe besuchte Schulform nicht identifizierbar. Die Schüler der gymnasialen Oberstufe rekrutieren sich aber überwiegend aus Besuchern der Gymnasialzüge der Sekundarstufe I (Cortina und Trommer 2003). Es ist deshalb insgesamt betrachtet zu erwarten, dass sich Schüler verschiedener sozioökonomischer Lagen hinsichtlich der Wahrscheinlichkeiten eines erfolgreichen Besuchs der gymnasialen Oberstufe nicht stark unterscheiden.

In Bezug auf die Ausfallproblematik haben die Analysen zum Auszug aus dem Elternhaus gezeigt, dass die Wahrscheinlichkeit auszuziehen steigt, wenn Schüler keine allgemein bildende Schule, sondern eine berufliche Schule besuchen oder studieren. Dies trifft insbesondere für Ostdeutschland zu. Dagegen unterschieden sich die Auszugsraten für Hochschulbesucher in Westdeutschland nur unwesentlich von denen der Schüler allgemein bildender Schulen. Zwischen der sozioökonomischen Lage der Herkunftsfamilie und dem Auszug konnten nur schwache Zusammenhänge ermittelt werden. Zusammenfassend deuten diese Befunde auf bedingt zufällige Ausfälle bei der Ausfallanalyse hin.

## **5.4.2 Statistische Modelle zu dem bis 1999 erreichten Abschluss**

Um Erfahrungen zu sammeln, wird zunächst auf die Berücksichtigung zeitabhängiger Ausfallmuster verzichtet. Der Zusammenhang zwischen dem Ausfall mit den einzelnen Statusübergängen des Schulbesuchs bzw. Bildungsabschlusses von 1996 bis 1999 wird anschließend untersucht. Die folgenden Analysen beschränken sich auf die Gruppe der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe im Jahre 1996 und ihr bis 1999 ermittelbares Bildungsergebnis (siehe hierzu Abb. 5.2 - 5.5; S. 78ff.).

Die nach 1996 zugezogenen Schüler und temporäre Ausfälle werden aus zwei Gründen nicht berücksichtigt. Zwar ist anzunehmen, dass Zugezogene und Fortgezogene gleiche Eigenschaften aufweisen, jedoch fehlen bei den zugezogenen Personen mit Abitur die Informationen darüber, ob sie im Jahr zuvor die gymnasiale Oberstufe besucht haben. Bedingt man auf den Besuch der gymnasialen Oberstufe im Zuzugsjahr, sind die aus dieser Auswahl resultierenden Zellenbeset-

zungen für differenzierte Analysen zu klein. Aus diesen Gründen muss auf die Informationen von Zugezogenen verzichtet werden.

Um schwache Zellenbesetzungen zu vermeiden, werden außerdem in der abhängigen Variablen (Status 1999;  $Y$ ) nur drei Kategorien unterschieden: (1) Besucher der gymnasialen Oberstufe, (2) Besucher einer beruflichen Schule und Personen ohne Schulbesuch, die als „Abgänger“ von der gymnasialen Oberstufe zusammengefasst werden, sowie (3) Absolventen mit Fachhochschulreife (FHR) oder allgemeiner Hochschulreife (AHR), denen auch die Besucher einer Fachhochschule oder Hochschule ohne Angaben zum erreichten allgemein bildenden Schulabschluss zugerechnet werden. Die Gruppe der Absolventen umfasst dabei auch Personen, die nach dem festgestellten Abschluss umgezogen sind.<sup>42</sup> Für gültige Werte in  $Y$  ist der Ausfallindikator  $R$  gleich Eins. Die bis 1999 weggezogenen Schüler, bei denen kein Abschluss ermittelt werden konnte, werden im Ausfallindikator mit  $R = 0$  codiert.

In inhaltlicher Hinsicht ist zu beachten, dass aus Fallzahlgründen für die statistische Erklärung des bis 1999 erreichten Status nur wenige Indikatoren herangezogen werden können. Die Bundesländer (Variable  $L$ ) werden nach der Regelschulbesuchszeit unterschieden. Bei den Alterskohorten ( $A$ ; Alter zum Zeitpunkt 1996) werden 18- und 19-Jährige zusammengefasst. Die sozioökonomische Lage des Elternhauses wird durch den Bildungsstatus ( $B$ )<sup>43</sup> und die sozialrechtliche Stellung im Beruf ( $S$ )<sup>44</sup> des Familienvorstands (i. d. R. des Vaters) abgegrenzt. Da diese Informationen zur sozialen Herkunft und die anderen Variablen auch mit dem Ausfall korreliert sein können, werden sie sowohl im Struktur- als auch im Ausfallmodell berücksichtigt. Zunächst folgen einfache Datenbeschreibungen.

Tabelle 5.6 zeigt Randverteilungen der ausgewählten Variablen. Wie bereits oben zu sehen war (siehe Abb. 5.2 - 5.5), werden in den vier Bundesländern mit zwölfjähriger Regelschulzeit bis zum Abitur um rund zehn Prozentpunkte höhere Abschlussquoten als in den anderen Bundesländern ermittelt. Mit rund 84 Prozent erreichen 18- bis 19-jährige Schüler wesentlich häufiger den Abschluss als jüngere Schüler. In Bezug auf die Ausfallquoten unterscheiden sich beide Variablen nur gering.

Eine geringe Variation der Ausfallquoten ist ebenfalls für das Bildungsniveau des Familienvorstands festzustellen. Jedoch unterscheiden sich die Ausfälle deutlich nach der beruflichen Stellung des Familienvorstands. Mit einem Anteil von 28 bis 29 Prozent fallen Jugendliche aus Arbeiterfamilien und von Nichterwerbstätigen vor einem ermittelbaren Abschluss stärker aus als Kinder von Selbstständigen (10,8 %) oder Beamten (15,1 %). Zugleich liegt der Anteil ermittelter Abschlüsse bei Arbeiterkindern mit rund 58 Prozent niedriger als bei Beamtenkindern (79,2 %). Ähnliche herkunftsspezifische Unterschiede – sowohl in Bezug auf die berufliche Stellung als auch das Bildungsniveau des Familienvorstands – zeigen sich bei den Anteilen der vom Gymnasium abgegangenen Schüler.

42 Nach Erreichen bzw. Ermittlung des Abschlusses sind 13,8 Prozent der Absolventen der hier verwendeten Teilstichprobe fortgezogen.

43 Fehlende Angaben zum Bildungsabschluss bei 63 Familienvorständen sind mit der Kategorie „Hauptschule oder weniger“ zusammengefasst worden. Dabei ist gewährleistet, dass diese Gruppe in Bezug auf die abhängige Variable ähnliche Verteilungen aufweist. Zu den mittleren Abschlüssen zählen Mittlere Reife, Polytechnische Oberschule, Fachhochschulreife und Abitur. Fachhochschul- und Hochschulabschluss werden in einer Bildungsgruppe zusammengefasst.

44 Aus den weiteren Analysen werden 29 Schüler ausgeschlossen, die nicht mehr bei den Eltern leben, so dass sich die Fallzahl gegenüber den in Abb. 5.2-5.5 dargestellten Auswertungen reduziert.

Tabelle 5.6: Bis 1999 erreichter Bildungsstatus der 1996 16- bis 19-jährigen Besucher der gymnasialen Oberstufe (Zeilenprozentwerte)

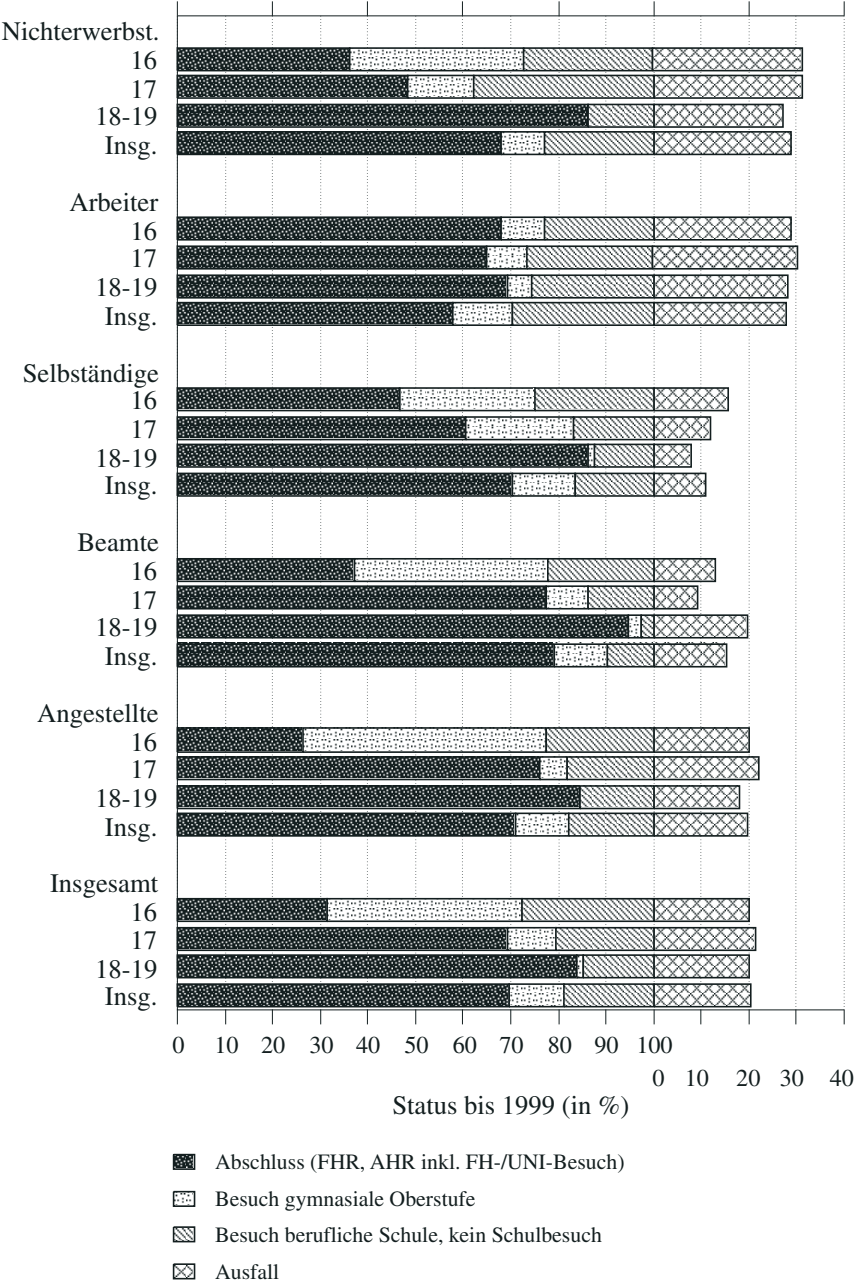
Variablen	Insgesamt (n)	Ohne Ausfall (R=1)			Mit Ausfall	
		Insges. (R=1)	darunter GYM	Abgang	Abschluss	(R=0)
Land/Regelschulzeit (L)						
13 Jahre/West	1.102	80,0	12,6	19,5	67,9	20,0
12 Jahre/Ost	207	77,3	6,3	16,2	77,5	22,7
Alterskohorte 1996 (A)						
16 Jahre	244	79,9	41,0	27,7	31,3	20,1
17 Jahre	414	78,7	10,1	20,6	69,3	21,3
18-19 Jahre	651	80,0	1,5	14,8	83,7	20,0
Berufliche Stellung FV (S)						
Nichterwerbstätige	128	71,1	8,8	23,1	68,1	28,9
Arbeiter	276	72,1	12,6	29,6	57,8	27,9
Selbstständige	185	89,2	13,3	16,4	70,3	10,8
Beamte	192	84,9	11,0	9,8	79,2	15,1
Angestellte	528	80,3	11,3	17,7	71,0	19,7
Bildungsniveau FV (B)						
Hauptschule, k.A.	462	79,4	11,2	28,3	60,5	20,6
Mittlere Abschlüsse	446	80,0	13,7	17,7	68,6	20,0
FH, Hochschule	401	79,3	9,7	9,8	80,5	20,7
Insgesamt	1.309	79,6	11,6	19,0	69,4	20,4

GYM: Gymnasiale Oberstufe.

Quelle: Mikrozensuspanel 1996-1999; eigene Berechnungen (ungewichtete Fallzahlen).

Diese ersten deskriptiven Befunde scheinen der oben skizzierten Hypothese schwacher Herkunftseffekte beim Abschluss der gymnasialen Oberstufe zu widersprechen. Aufgrund fehlender Kontrollvariablen und der bisher nicht geklärten Frage selektiver Ausfälle sowie der eingeschränkten Qualität der Bildungsangaben sollte allerdings dieses Zwischenergebnis nicht überbewertet werden.

Die nach beruflicher Stellung des Familienvorstands und Alterskohorte differenzierten Verteilungen vermitteln weitere Aufschlüsse zum Zusammenhang zwischen Schulerfolg und Ausfall. Abbildung 5.9 zeigt für jede berufliche Stellung einen mit dem Alter korrelierten Anteil erfolgreicher Schulabschlüsse.



Die Anteile für Ausfall beziehen sich auf alle Personen; die anderen Anteile beziehen sich nur auf räumlich Immobiler.

Quelle: Mikrozensuspanel 1996-1999; eigene Berechnungen (siehe Tab. 5.6).

Abbildung 5.9: Bis 1999 erreichter Bildungsstatus der Besucher der gymnasialen Oberstufe des Jahres 1996 nach beruflicher Stellung des Familienvorstands und Alter

Schüler aus Familien von Nichterwerbstätigen und Arbeitern unterscheiden sich sowohl durch einen überdurchschnittlichen Anteil bei den Abgängern der gymnasialen Oberstufe (Berufliche Schule/ohne Schulbesuch 1999) als auch bei den Ausfällen von den anderen Herkunftsgruppen, insbesondere Beamten.

Unter der Annahme eines engen Zusammenhanges zwischen Ausfall und Abschluss ist zu erwarten, dass die kohortenspezifische Ausfallrate korrespondierend mit dem Anteil der Abschlüsse steigt. Betrachtet man dagegen die nach Kohorten differenzierten Gesamtverteilungen, zeigt sich eine im Wesentlichen konstante Ausfallrate. Nur für die Gruppe der Beamtenkinder trifft der erwartete Zusammenhang näherungsweise zu.

Für die folgenden Modelle deuten diese vorläufigen Ergebnisse auf starke Zusammenhänge zwischen Alterskohorte und Status sowie schwache Zusammenhänge zwischen Ausfall und Alter hin. Zwischen Ausfall und beruflicher Stellung des Familienvorstands sind die Zusammenhänge dagegen stärker.

Tabelle 5.7 zeigt die Ergebnisse von log-linearen Modellen zum Ausfall. Der Einfachheit halber wird nur die Logit-Spezifikation genannt. Es werden zunächst verschiedene Modelle unter der Annahme eines weder mit der abhängigen noch mit unabhängigen Variablen zusammenhängenden Ausfalls (MCAR) berichtet. Die Ergebnisse des Strukturmodells sind mit Analysen identisch, die unter Ausschluss der Ausfälle durchgeführt werden. Die Modellvergleiche 2 vs. 1 und 7 vs. 6 bestätigen den bereits auf Basis der deskriptiven Analyse gefundenen engen Zusammenhang zwischen Status und Alterskohorte ( $YA$ ). Sowohl der durch das Pseudo- $R^2$  ermittelte Bruttoeffekt (Modell 2 vs. 1) als auch der Nettoeffekt (Modell 7 vs. 6) des Alters ( $A$ ) beträgt jeweils rund 14 Prozent.<sup>45</sup> Die Erklärungskraft der anderen Variablen ist mit Abstand geringer. Der statistisch vergleichsweise schwache Zusammenhang der Indikatoren der sozialen Herkunft mit dem bis 1999 ermittelbaren Schulbesuchs- bzw. Bildungsstatus korrespondiert mit den oben skizzierten Annahmen. Der partielle Erklärungsbeitrag der Variable Stellung im Beruf des Familienvorstands ( $S$ ) ist statistisch nicht signifikant (Modell 9 vs. 6).

---

45 Das Pseudo- $R^2$  eines gegebenen Logit-Modells  $M$  ist:  $R^2 = 1 - (\text{Log-Likelihood} / \text{Log-Likelihood } M_0)$ , wobei  $M_0$  das Modell bezeichnet, das nur die Regressionskonstanten enthält. Das Pseudo- $R^2$  ist im Vergleich zum  $R^2$  der linearen Regression i. d. R. wesentlich niedriger. Werte über  $R^2 = 0,2$  sprechen für eine außerordentlich gute Modellanpassung (Hensher und Johnson 1981: 51).

Tabelle 5.7: Log-lineare Modelle zum Ausfall und Status bis 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996

#	Logit-Spezifikation	Likelihood-Ratio-Statistik						R <sup>2</sup> (in %)	
		Log-Likelihood	Param.	G <sup>2</sup>	df	P	AIC	Struktur-mod.	Ausfall-mod.
MCAR-Modelle									
1	(Y) (R)	-5.845,51	3				11.697	0	
2	#1 + YA	-5.726,11	7	238,80	4	0,00	11.466	14,0	
3	#1 + YL	-5.841,51	5	7,99	2	0,02	11.693	0,5	
4	#1 + YS	-5.831,29	11	28,44	8	0,00	11.685	1,7	
5	#1 + YB	-5.823,22	7	44,58	4	0,00	11.660	2,6	
6	#1 + YA, YL, YS, YB	-5.688,10	21	314,82	17	0,00	11.418	18,4	
7	#6 - YA	-5.811,55	17	246,90	4	0,00	11.657	4,0	
8	#6 - YL	-5.700,29	19	24,39	2	0,00	11.439	17,0	
9	#6 - YS	-5.695,55	13	14,90	8	0,06	11.417	17,6	
10	#6 - YB	-5.699,44	17	22,69	4	0,00	11.433	17,1	
MAR-Modelle									
11	#6 + RA	-5.687,96	23	0,28	2	0,87	11.422	18,4	0,0
12	#6 + RL	-5.687,70	22	0,79	1	0,37	11.419	18,4	0,1
13	#6 + RS	-5.673,14	25	29,92	4	0,00	11.396	18,4	2,3
14	#6 + RB	-5.688,06	23	0,09	2	0,96	11.422	18,4	0,0
15	#6 + RA, RL, RS, RB	-5.670,16	30	35,89	9	0,00	11.400	18,4	2,7
NINR-Modell									
16	#13 + RY	-5.672,80	27	0,68	2	0,71	11.400	18,8	4,8

Unter den Modellen zur Annahme bedingt zufälliger Ausfälle (MAR) werden ausgehend von Modell 6 Zusammenhänge zwischen dem Ausfallindikator und den erklärenden Variablen geschätzt.<sup>46</sup> Hierbei zeigt sich, dass der Ausfall sehr stark mit der Variablen Stellung im Beruf (S) zusammenhängt und auf die Interaktionen mit den anderen Variablen praktisch verzichtet werden könnte, da sich das Modell 15 nicht signifikant vom Modell 13 unterscheidet ( $G^2 = 5,97$ ; d.f. = 5;  $P = 0,31$ ). Zu beachten ist, dass für die Ausfallmodelle die durch das Pseudo-R<sup>2</sup> gemessene statistische Erklärungsleistung nur bei 2,7 Prozent liegt, d. h. die verwendeten Variablen insgesamt nicht hoch mit dem Ausfall korrelieren.

Ob über den Zusammenhang zwischen der Stellung im Beruf des Familienvorstands (S) und dem Ausfall hinausgehend ein nicht ignorierbarer Ausfall vorliegt, wird mit dem NINR-Modell 16 geprüft. Zwar steigt das für diese Teiltabelle ermittelte Pseudo-R<sup>2</sup> auf fast fünf Prozent an, doch ist die für die Gesamtdaten ermittelte statistische Erklärung dieses Modells nicht signifikant vom Modell 13 verschieden. Die Modellergebnisse sprechen somit für die Annahme eines Ausfalls, der hauptsächlich mit der Variablen Stellung im Beruf des Familienvorstands (S) zusammenhängt. Sie passen zu den deskriptiven Befunden (siehe Tab. 5.6 und Abb. 5.9). Die für die Ausfallgruppe durch das Modell 13 geschätzten Anteile entsprechen den Verteilungen in Tabelle

46 Weil ausschließlich diskrete Daten analysiert werden, zählen streng genommen die MAR-Modelle bereits zu den NINR-Modellen (siehe Seite 28).



5.6. Gleichwohl kann die Akzeptanz der MAR-Annahme nur als ein vorläufiges Ergebnis zählen, da bisher keine Angaben über den Zeitpunkt des Ausfalls bzw. Wegzugs berücksichtigt wurden.

### 5.4.3 Statistische Modelle für einzelne Statusübergänge

Mithilfe weiterer Informationen zum Ausfall kann die Schätzung der Ausfallwahrscheinlichkeiten verbessert werden. Insbesondere ermöglicht die Verbleibsdauer im Panel einen differenzierteren Vergleich zwischen räumlich mobilen und immobilen Schulbesuchern. Im Unterschied zu den obigen Modellen kann bei der Berücksichtigung einzelner Statusübergänge nicht nur die MAR-Annahme genauer untersucht werden, in welcher Weise der Ausfall mit soziodemografischen Variablen verbunden ist, sondern auch, ob der Ausfall mit dem davor eingenommenen nicht beobachteten Status in Beziehung steht.

Um zu schwache Zellenbesetzungen zu vermeiden, wird die Variable Bildungsniveau des Familienvorstands ( $B$ ) nicht weiter berücksichtigt. In Bezug auf Statusübergänge treten strukturelle Nullzellen auf. Insbesondere ist bei der Kategorie „Abschluss“ zu beachten, dass nachfolgende Übergänge zum Besuch der gymnasialen Oberstufe oder Wechsel in die Kategorie „Abgang von der Oberstufe“ (Berufliche Schule/kein Schulbesuch) nicht möglich sind. Gleichfalls sind für Abgänger Wechsel zum Status „Abschluss“ und „Schulbesuch“ ausgeschlossen.<sup>47</sup> Im Strukturmodell werden aus diesen Gründen die Übergänge vom Besuch der gymnasialen Oberstufe zum Abgang und zum Abschluss modelliert. Von diesen beiden Zuständen aus kann im Zeitverlauf infolge eines Wegzugs allerdings ein Wechsel zum Ausfall erfolgen.

Analog zum Strukturmodell lassen sich im Ausfallmodell ebenfalls Übergangswahrscheinlichkeiten (z. B.:  $R_t R_{t-1}$  oder  $R_t Y_{t-1}$ ) schätzen (vgl. Conaway 1992, 1993). Da die in der Analyse herangezogenen Daten keine nach 1996 zugezogenen und temporär ausgefallenen Personen enthalten, liegen so genannte geschachtelte oder monotone Ausfallmuster vor. Bei der Modellierung sind die logisch unmöglichen Übergänge (z. B. von „Ausfall 1997“ zu „Teilnahme 1998“) als strukturelle Nullzellen zu behandeln.

Selbst mit den wenigen für die Analyse herangezogenen Variablen können für die Übergänge 1996 bis 1999 eine Vielzahl von Modellen geschätzt werden. Tabelle 5.8 enthält die Ergebnisse für einige unter inhaltlichen und methodischen Aspekten ausgewählte Modelle. Den Startpunkt der Analysen stellt das MCAR-Modell 1 dar, das im Strukturteil die Interaktionen zwischen den Statusvariablen  $Y_t$  mit den drei Analysevariablen Alterskohorte ( $A$ ), Bundesland ( $L$ ) und Stellung im Beruf des Familienvorstands ( $S$ ) enthält. Darüber hinaus werden in diesem Modell mit den Regressionskoeffizienten zu  $Y_t Y_{t-1}$  die Übergangswahrscheinlichkeiten von der gymnasialen Oberstufe zu Abgängen und Abschlüssen geschätzt. Die Modelle 2 bis 4, in denen die Interaktionen jeweils einer Analysevariablen aus Modell 1 herausgenommen sind, zeigen, dass das Alter ( $A$ ) am stärksten mit den Statusvariablen  $Y_t$  zusammenhängt.

47 Der Abgang von der gymnasialen Oberstufe und spätere Wiedereintritt ist zwar prinzipiell möglich. Entsprechende Übergangsmuster finden sich auch in den Daten, sie lassen sich aber nicht klar von inkonsistenten Antworten abgrenzen. Da es sich um insgesamt nur 58 Schüler handelt, wurden solche Angaben nach Dateninspektionen in Übergangsmuster recodiert, in denen Wechsel von Abgang zu Schulbesuch oder Abschluss nicht mehr vorkommen.



Tabelle 5.8: Log-lineare Modelle zum Ausfall und Status 1997 bis 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996

#	Logit-Spezifikation	Likelihood-Ratio-Statistik					R <sup>2</sup> (in %)	
		Log-Likelihood	Param.	G <sup>2</sup>	df	P	AIC	Ausfall-mod.
MCAR-Modelle								
1	$Y_i^*(A, L, S, Y_{i-1}), R$	-5.426,92	80				11.014	
2	#1 - $Y_i A$	-5.742,39	68	630,95	12	0,00	11.621	
3	#1 - $Y_i L$	-5.471,06	74	88,28	6	0,00	11.090	
4	#1 - $Y_i S$	-5.459,15	56	64,46	24	0,00	11.030	
MAR-Modelle								
5	#1 + $R_i A$	-5.415,84	86	22,16	6	0,00	11.004	0,9
6	#1 + $R_i L$	-5.422,37	83	9,10	3	0,03	11.011	0,4
7	#1 + $R_i S$	-5.407,63	92	38,58	12	0,00	10.999	1,6
8	#1 + $R_i A, R_i L, R_i S$	-5.391,75	101	70,33	21	0,00	10.986	3,0
9	#8 - $R_i A$	-5.403,49	95	23,48	6	0,00	10.997	2,0
10	#8 - $R_i L$	-5.396,94	98	10,38	3	0,02	10.990	2,6
11	#8 - $R_i S$	-5.409,92	89	36,34	12	0,00	10.998	1,5
NINR-Modelle								
12	#8 + $R_i Y_{i-1}$	-5.384,22	105	15,07	4	0,00	10.978	4,1
13	#8 + $R_{98} Y_{98}, R_{99} Y_{99}$	-5.383,96	105	15,58	4	0,00	10.978	5,1

Die auf Modell 1 aufbauenden MAR-Modelle<sup>48</sup> ermitteln zunächst die Bruttoeffekte des Zusammenhangs zwischen dem Ausfall und den unabhängigen Variablen (Modelle 5-7 vs. Modell 1). Wie anhand des Pseudo-R<sup>2</sup> für die Ausfallgruppe zu erkennen, ist der Bruttoeffekt der Variablen Stellung im Beruf des Familienvorstands ( $S$ ) am stärksten.<sup>49</sup> Dies gilt auch in Bezug auf die Nettoeffekte (Modelle 9-11 vs. Modell 8). Das partielle Pseudo-R<sup>2</sup> der Variablen Stellung im Beruf beträgt rund 1,5 Prozent, während der partielle Effekt der Variablen Bundesland ( $L$ ) bei einem Prozent und der Effekt der Kohorte ( $A$ ) bei nur 0,4 Prozent liegen.

Unter der Annahme, dass der Wegzug mit der Situation vor dem Ausfall in Beziehung steht, werden in Modell 12 Interaktionen zwischen dem Ausfall ( $R_i$ ) und dem Status vor dem Ausfall ( $Y_{i-1}$ ) geschätzt. Hierbei handelt es sich um ein Modell zum nicht ignorierbaren Ausfall (NINR), da beispielsweise für Schüler, die 1997 ausgefallen sind, der Schulbesuchsstatus 1998 nicht beobachtet wird. Gegenüber dem MAR-Modell 8 kann mit dem Modell 12 eine zwar eher schwache, aber signifikant größere statistische Erklärung erreicht werden.

48 Auch hier gilt die auf Seite 28 genannte Einschränkung zur Unterscheidung zwischen MAR- und NINR-Modellen. Da die Koeffizienten zu den inhaltlich interessierenden Analysevariablen des Strukturmodells bei den verschiedenen Modellen weitgehend stabil sind (siehe Tabelle A.3), werden die idealtypischen Modellbezeichnungen dennoch verwendet.

49 Aus Platzgründen werden für das Strukturmodell keine Pseudo-R<sup>2</sup>-Werte berichtet. Siehe dazu für ausgewählte Modelle Tabelle A.3 im Anhang. Die im Zeitverlauf beachtlich ansteigende statistische Erklärung ist insbesondere auf die Modellierung der Übergangswahrscheinlichkeiten  $Y_i Y_{i-1}$  zurückzuführen.

Im nächsten Schritt wird modelliert, dass der Ausfall mit dem nicht beobachteten Status zu dem gegebenen Zeitpunkt ( $R_i Y_i$ ) zusammenhängt. Aufgrund von Identifikationsproblemen bzw. einer „boundary solution“ der Parameter für den Zeitpunkt 1997 werden mit dem Modell 13 nur die Koeffizienten zu  $R_{98} Y_{98}$  und  $R_{99} Y_{99}$  geschätzt (siehe zur „boundary solution“ Seite 34). Mit diesem Modell lässt sich die statistische Erklärung im Vergleich zum MAR-Modell 8 weiter verbessern. Die statistischen Kennziffern der Modelle 12 und 13 sind aber praktisch gleich. Diese Ergebnisse sprechen zunächst für die Annahme selektiver Ausfälle bzw. Umzüge in Abhängigkeit vom (früheren bzw. gegenwärtigen) Status. Sie widersprechen damit den im Unterabschnitt 5.4.2 gezogenen vorläufigen Schlussfolgerungen.

Orientiert man sich bei der Modellwahl zunächst nur an den AIC-Werten, kommen sowohl die beiden NINR-Modelle 12 und 13 als auch – mit Abstand – das MAR-Modell 8 als die Modelle mit den besten, d. h. niedrigsten AIC-Werten in Frage. Diese Entscheidung ist mit weitreichenden Konsequenzen verbunden. Das MAR-Modell impliziert, dass die infolge von Ausfällen fehlenden Werte die Schätzung der Zusammenhänge zwischen den Analysevariablen ( $Y, X$ ) im Strukturmodell nicht beeinträchtigen, sofern entsprechende Kontrollvariablen im Modell enthalten sind. Unter der MAR-Annahme kann man mit dem Kehrwert der Wahrscheinlichkeit für die Immobilität gewichten und erhält damit konsistente Schätzungen. Beim NINR-Modell wird aber davon ausgegangen, dass der Ausfall vom (nicht beobachteten) Status  $Y$  abhängt. Entsprechend der Ausgangshypothese träfe dies zu, wenn Oberstufenschüler nach dem Abitur umziehen bevor der Abschluss im Mikrozensus durch den Vergleich der Angaben der Erhebungszeitpunkte festgestellt werden kann. Unter NINR-Annahme müssen, wie in den obigen Modellen, der Ausfall und das jeweils interessierende inhaltliche Modell simultan geschätzt werden.

Mithilfe der unter den jeweiligen Modellannahmen geschätzten Anteile der abhängigen Variablen ist es ansatzweise möglich, einzuschätzen, wie plausibel diese völlig verschiedenen Annahmen sind. Tabelle 5.9 enthält die zum Zeitpunkt 1999 geschätzten Anteile der abhängigen Variablen für die räumlich mobilen Schüler, die sich bis 1999 auf 28 Prozent des Anfangsbestandes aufsummieren. Die durch das Modell 4 unter der Annahme eines rein zufälligen Ausfalls (MCAR) geschätzten Anteile entsprechen im Wesentlichen den für die Immobilen geschätzten Anteilen. Das MCAR-Modell 4 bietet sich deshalb als Vergleichsbasis an. Unter der Annahme, dass die Ausfälle zwar mit den unabhängigen Variablen, aber nicht mit dem Schulbesuchs- bzw. Abschlussstatus ( $Y_i$ ) verbunden sind, werden im MAR-Modell 8 ein niedrigerer Anteil von Schulabgängern (17,5 %), deutlich niedrigere Anteile des Schulbesuchs (7,9 %) und ein etwas höherer Anteil der Absolventen (74,6 %) geschätzt. Im Unterschied zum MAR-Modell 8 werden durch die Modelle 12 und 13, die zusätzlich zur MAR-Annahme einen nicht ignorierbaren Ausfall unterstellen, weniger Oberstufenschüler und Absolventen, dagegen aber erheblich mehr Schulabgänger geschätzt.

Tabelle 5.9: Geschätzter Bildungsstatus im Jahre 1999 verschiedener Modelle nach der Stellung im Beruf des Familienvorstands für die Ausfallgruppe – Spaltenprozentwerte

Status 1999	Stellung im Beruf des Familienvorstands					
	Insgesamt	Nichter- werbstätige	Arbeiter	Selbst- ständige	Beamte	Angestellte
<i>MCAR-Modell 1</i>						
Gymnasiale Oberstufe	10,5	8,0	10,1	13,0	10,5	10,5
Abgang	18,1	22,5	29,9	12,8	8,7	16,2
Abschluss	71,4	69,6	60,0	74,2	80,8	73,3
<i>MAR-Modell 8</i>						
Gymnasiale Oberstufe	7,9	6,3	8,0	9,6	8,0	8,0
Abgang	17,5	20,7	27,7	11,4	7,8	14,7
Abschluss	74,6	73,0	64,4	79,1	84,2	77,4
<i>NINR-Modell 12</i>						
Gymnasiale Oberstufe	6,5	5,1	6,1	8,4	7,0	6,7
Abgang	24,2	28,5	37,8	15,3	10,6	20,7
Abschluss	69,2	66,4	56,1	76,3	82,4	72,6
<i>NINR-Modell 13</i>						
Gymnasiale Oberstufe	7,2	5,3	6,6	9,9	7,6	7,5
Abgang	27,7	34,0	42,2	17,1	12,8	23,8
Abschluss	65,1	60,8	51,2	73,0	79,5	68,7

Die beiden NINR-Modelle unterscheiden sich vorrangig hinsichtlich des Absolventenanteils, der mit Modell 13 niedriger als mit Modell 12 geschätzt wird. Diese Unterschiede hängen mit den verschiedenen Annahmen über die Ausfälle zusammen. Während im NINR-Modell 12 ein Zusammenhang mit dem vorherigen Status ( $R_t Y_{t-1}$ ) angenommen wird, geht das NINR-Modell 13 davon aus, dass der Ausfall mit dem zeitgleichen (nicht beobachteten) Status zusammenhängt ( $R_t Y_t$ ).

Differenziert man die Schätzungen in Tabelle 5.9 nach der beruflichen Stellung des Familienvorstands, steigen die Abschlussquoten des MAR-Modells gegenüber dem MCAR-Modell zulasten der Anteile für den Schulbesuch und für die Abgänger. Die durch das MAR-Modell geschätzten Anteile liegen mit bis zu fünf Prozentpunkten Unterschied vergleichsweise nahe bei den MCAR-Schätzungen. Hinsichtlich der durch die NINR-Modelle geschätzten Anteile sind jedoch für die Gruppen der Nichterwerbstätigen und Arbeiter für den Abgang und Abschluss größere Prozentsatzdifferenzen im Vergleich zum MCAR-Modell festzustellen. Während zum Beispiel mit dem MCAR-Modell für Arbeiterkinder der Anteil der Oberstufenabgänger auf rund 30 Prozent geschätzt wird, liegen diese Werte im Modell 12 bzw. Modell 13 mit 38 bzw. 42 Prozent um sechs bzw. zwölf Prozentpunkte höher. Dies erscheint plausibel, da in der deskriptiven Analyse (siehe Tab. 5.6 und Abb. 5.9) bei Kindern von Arbeitern und Nichterwerbstätigen sowohl überdurchschnittliche Ausfälle als auch höhere Abgängerquoten festgestellt wurden. Dennoch soll im Folgenden weiter darüber nachgedacht werden, weshalb gerade die Variable Stellung im Beruf des Familienvorstands so eng mit dem Ausfallgeschehen interagiert. Dabei ist auch zu beachten, dass die im vierten Kapitel durchgeführten Analysen sich auf Auszüge aus dem Elternhaus konzentrierten und die räumliche Mobilität ganzer Haushalte ausgeblendet werden musste.

Die Abgänger der Oberstufe setzen sich aus Besuchern beruflicher Schulen und Personen ohne Schulbesuch zusammen. Diese Jugendlichen verfügen i. d. R. über ein höheres eigenes Einkommen als Jugendliche, die noch die Oberstufe besuchen. Der enge Zusammenhang zwischen Oberstufenabgang und Ausfall wäre dann damit zu erklären, dass diese Gruppe aufgrund der größeren materiellen Unabhängigkeit von den Eltern eher von zu Hause auszieht. Untersucht man aber die herkunftsspezifischen Auszugsmuster für die Panelausfälle, zeigt sich auf der Personenebene eine unterdurchschnittliche Rate des Auszugs aus dem (weiter befragten) Elternhaushalt für Kinder von Arbeitern (51 % vs. 62 % insgesamt). Dagegen ist auf der Haushaltsebene für Kinder von Arbeitern mit 49 Prozent eine überdurchschnittliche Haushaltsmobilität (38 % insgesamt) festzustellen. Wenn aber die Ausfälle der Arbeiterkinder mit einer höheren Umzugswahrscheinlichkeit des gesamten Haushalts zusammenhängen, dann spricht dies insgesamt eher für die Annahme bedingt zufälliger Ausfälle (MAR) in Abhängigkeit von der Stellung im Beruf des Familienvorstands.

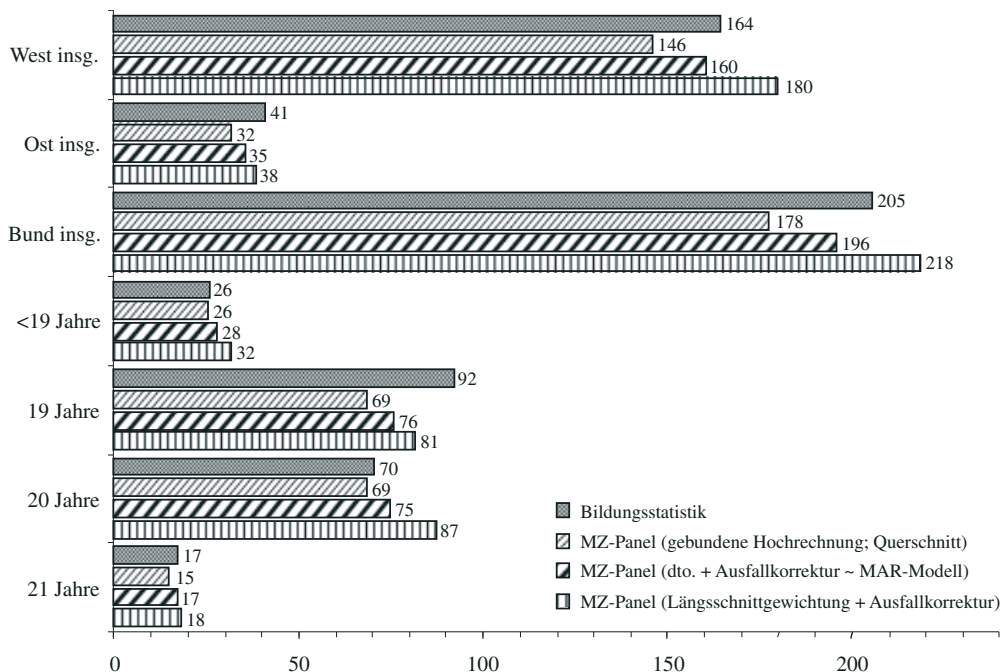
Es wäre jedoch nicht sachgerecht, ein Modell und die damit verbundene Annahme zum Ausfallgeschehen ausschließlich aufgrund des AIC-Kriteriums und den Plausibilitätsüberlegungen zu den geschätzten Anteilen zu präferieren. Wie oben angesprochen, können diese statistischen Verfahren dabei helfen, darüber nachzudenken, wie die Ausfälle zustande gekommen sind und gegebenenfalls Korrekturverfahren für die fehlenden Daten zu entwickeln. Die Modellergebnisse selbst sind allerdings von den für die Gruppe ohne Ausfälle vorliegenden Daten und den zugrunde liegenden Annahmen, hier insbesondere dem Ausschluss höherer Interaktionen, abhängig. Des Weiteren gehen die Modelle von fehlerfreien Daten aus. In Bezug auf die festgestellten Inkonsistenzen der Bildungsangaben ist dies eine kritische Annahme. Darüber hinaus lassen sich die Modellannahmen lediglich auf Basis der verwendeten Daten grundsätzlich nicht prüfen (Allison 2002: 4-5). Da für die ausgefallenen Schüler keine Informationen zu ihrem erreichten Status vorliegen, kann die MAR-Annahme, dass sich die Gruppen der vor Ermittlung des Abschlusses weggezogenen und nicht weggezogenen Schüler hinsichtlich der beruflichen Stellung des Familienvorstands nicht systematisch unterscheiden, nicht kontrolliert werden. Zur Validierung der NINR-Modelle sind fundierte a priori Kenntnisse sowie externe Daten notwendig. Zu der hier untersuchten Fragestellung liegen jedoch, wie bereits erwähnt, solche Informationen nur sehr begrenzt vor.

Eine Möglichkeit, die Plausibilität der Modelle näherungsweise zu prüfen, ist durch den Vergleich der Ergebnisse des Mikrozensuspanels mit den Absolventenzahlen der Bildungsstatistik gegeben. Gilt die Annahme bedingt zufälligen Ausfalls (MAR), kann eine Ausfallkorrektur durch Gewichtung mit dem Kehrwert der Wahrscheinlichkeit für die Immobilität erreicht werden. Für die Schätzung dieser Wahrscheinlichkeiten wird im Folgenden die Variable Stellung im Beruf des Familienvorstands herangezogen, die in den Analysen die größte statistische Erklärungskraft aufweist.

Abbildung 5.10 zeigt, dass die Absolventen bei einfacher gebundener Hochrechnung unter Verwendung der Querschnittsgewichte für 1996<sup>50</sup> im Mikrozensuspanel im Vergleich zur Bildungsstatistik um 13 Prozent (177,7 tsd. vs. 205,4 tsd. Absolventen; hier ohne Darstellung) untererfasst werden. Ergänzt man diesen Gewichtungsfaktor mit dem aus einem Logit-Modell unter Verwendung der Variablen Ländergruppe (*L*) und Stellung im Beruf des Familienvorstands (*S*) ermittelten Kehrwert der Wahrscheinlichkeit für räumliche Immobilität, reduziert sich die Untererfassung auf weniger als fünf Prozent (195,7 tsd. vs. 205,4 tsd. Absolventen; siehe Abb. 5.10). In den Ländern mit zwölfjähriger Regelschulzeit kann eine Verringerung der Untererfassung von 23

50 Siehe zur Hochrechnung Fußnote 36, Seite 85.

auf 14 Prozent erreicht werden. Für Westdeutschland ist der Rückgang von elf auf zwei Prozent erheblich stärker. Konzentriert man sich beim Alter auf die beiden großen Gruppen, zeigt sich bei den 19-Jährigen mit einer besseren Anpassung der ausfallgewichteten Panelergebnisse an die Bildungsstatistik, aber einer Übererfassung bei den 20-Jährigen, ein gemischtes Ergebnis.



#### Quellen:

Bildungsstatistik: Statistisches Bundesamt (1997a,b). (Absolventen bis unter 21 Jahren in den Ländern mit 13-jährigem Schulbesuch n = 164,4 tsd., in den Ländern mit zwölfjährigem Schulbesuch n = 41,0 tsd. Abiturienten; ohne Schulfremdenprüfungen).

Mikrozensuspanel (ungewichtete Fallzahlen: n = 261).

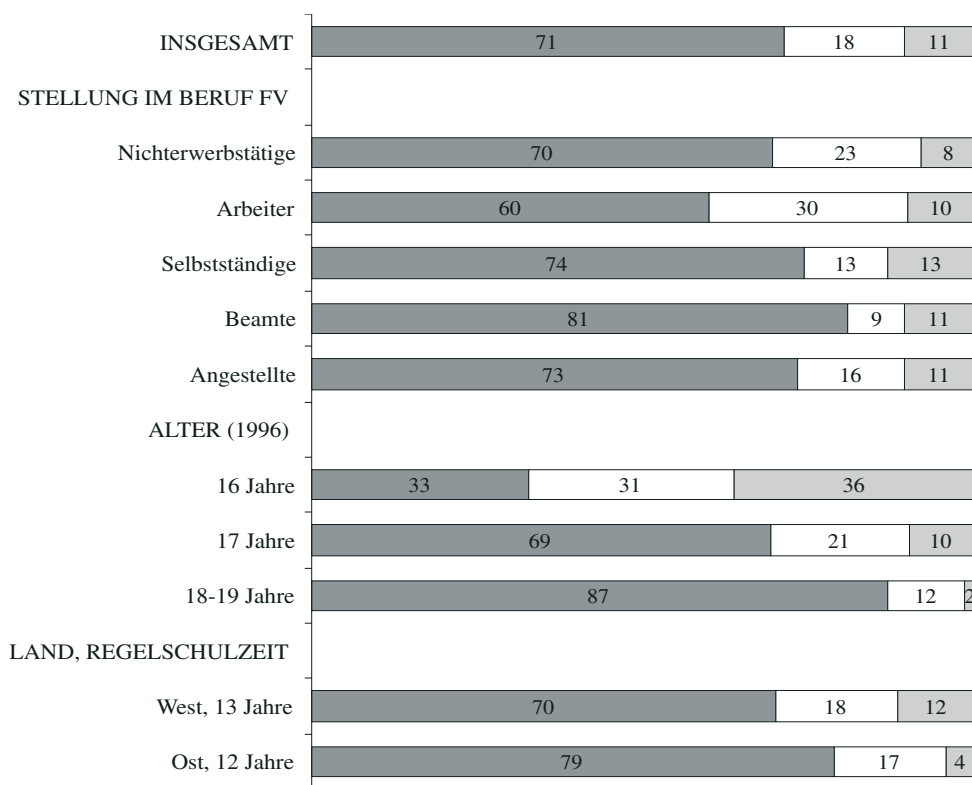
Abbildung 5.10: Absolventen allgemein bildender Schulen (bis 21 Jahren) mit Allgemeiner Hochschulreife 1996 nach Region und Alter (Geburtsjahr) (in 1.000)

Die Verwendung eines Längsschnittgewichts, in dem neben der Anpassung an demografische Randverteilungen für die Erhebungszeitpunkte 1996 und 1997 zusätzlich Logit-Schätzungen zur räumlichen Mobilität berücksichtigt werden (Basic et al. 2005; Rohloff 2005), führt zu einer Übererfassung der Abiturientenzahlen des Mikrozensuspanels (218,4 tsd. vs. 205,4 tsd. Absolventen in der Bildungsstatistik) und insgesamt etwas größeren Abweichung im Vergleich zum speziell konstruierten Korrekturgewicht.

In Bezug auf die Annahmen bedingt zufälliger vs. nicht ignorierbarer Ausfälle ist zu beachten, dass beide NINR-Modelle die Absolventenanteile für die ausgefallenen Schüler um fünf bzw. zehn Prozentpunkte geringer schätzen als das MAR-Modell. Folglich würden sich mit einer den NINR-Annahmen entsprechenden Ausfallgewichtung größere Abweichungen zur Bildungsstatistik ergeben als bei der hier vorgenommenen Gewichtung unter der MAR-Annahme. Dies spricht für die Annahme bedingt zufälliger Ausfälle. Akzeptiert man deshalb das MAR-Modell 8, darf aber nicht außer Acht bleiben, dass im Grunde genommen keine entscheidungsfähige Situa-

tion vorliegt, da die Daten des Mikrozensuspanels infolge von Fehlklassifikationen im Unterschied zur Bildungsstatistik auch Schüler beruflicher Schulen enthalten.

Unter diesem Vorbehalt zeigt Abbildung 5.11, welcher Bildungsstatus für die Gesamtdaten (einschl. Ausfällen) im Jahre 1999 für die 16- bis 19-jährigen Besucher der gymnasialen Oberstufe durch das MAR-Modell 8 geschätzt wird. Nur 71 Prozent der Schüler erreichen einen Abschluss. Ebenfalls sind die „Abschlussquoten“ für die 17-Jährigen, die 1996 überwiegend die elfte Klasse besucht haben dürften, unplausibel niedrig. Die Übergänge zum Besuch beruflicher Schulen oder keinem Schulbesuch betragen 18 Prozent, bei 16-Jährigen sogar 31 Prozent. Die Unterschiede zwischen den Ländern mit 13- und zwölfjähriger Regelschulzeit (West / Ost) dürften kaum auf den unterschiedlichen schulischen Erfolg, sondern vielmehr auf die in Westdeutschland vergleichsweise kürzere Beobachtungszeit im Panel zurückzuführen sein, da z. B. bei Klassenwiederholungen ein Abschluss erst nach 1999 ermittelbar wäre.



Bildungsstatus 1999 (MAR-Modell 8; siehe Tab. 5.8 und 5.6)

- Abschluss mit Allgemeiner Hochschulreife oder Fachhochschulreife)
- Abgang (Besuch einer beruflichen Schule oder kein Schulbesuch)
- ▒ Besuch der gymnasialen Oberstufe

Abbildung 5.11: Geschätzter Bildungsstatus im Jahre 1999 für die 1996 16- bis 19-jährigen Besucher der gymnasialen Oberstufe einschließlich Ausfällen (in Prozent)

Hinsichtlich den nach sozialer Herkunft differenzierten Chancen des erfolgreichen Abschlusses der gymnasialen Oberstufe wurde mit Bezug auf Theorien zur Bildungsungleichheit vermutet, dass die sozialen Ungleichheiten in der gymnasialen Oberstufe gering sein werden. Abbildung 5.11 zeigt dagegen für Arbeiterkinder mit 60 Prozent unterdurchschnittliche und vor allem erheblich geringere „Erfolgswahrscheinlichkeiten“ als für Kinder von Beamten (81 %). Dies spricht gegen die Hypothese. Andererseits ist das Chancenverhältnis für Arbeiterkinder von 1,35 (81 % / 60 %) zu Beamtenkindern aber wesentlich geringer als die sozialen Unterschiede des Gymnasialschulbesuchs in der Sekundarstufe I. So hatten nach Ergebnissen des Mikrozensus 1989 (Westdeutschland), mit dem letztmals die Schulformen des dreigliedrigen Schulwesens erfasst wurden, 13- bis 15-jährige Beamtenkinder eine rund fünfmal größere Chance des Gymnasialschulbesuchs als Arbeiterkinder (hier ohne Nachweis; vgl. Schimpl-Neimanns 2000). Mit einer anderen, allerdings nicht vergleichbaren Klassifikation wurden in der ersten Pisa-Studie im Jahre 2000 die sozialen Unterschiede des Gymnasialschulbesuchs 15-Jähriger (Sekundarstufe I) zwischen Kindern aus der oberen bzw. unteren Dienstklasse und von Facharbeitern ein Chancenverhältnis von 4,3 bzw. 3,3 (Baumert und Schümer 2001: 357) ermittelt. Dennoch sollten die mit dem Mikrozensuspanel durchgeführten Schätzungen aufgrund der systematischen Fehler der Erfassung der Sekundarstufe I vs. II und der gymnasialen Oberstufe vs. berufliche Schulen (siehe Unterabschnitt 5.3.3) nicht in Bezug auf soziale Ungleichheiten des Bildungserfolgs in der gymnasialen Oberstufe interpretiert werden.

## 5.5 Zusammenfassung

In diesem Kapitel wurden erstmals Verläufe vom Besuch der gymnasialen Oberstufe bis zum Abschluss mit dem Abitur mit Mikrozensusdaten untersucht. Bei der Prüfung, ob diese Bildungsverläufe u. a. zur Ermittlung von Quoten des erfolgreichen Abschlusses der gymnasialen Oberstufe mit dem Abitur eingesetzt werden können, ergaben sich allerdings unplausibel niedrige Abschlussquoten.

Um den Gründen für die niedrige Abiturientenquote im Panel nachzugehen, wurden deshalb Ergebnisse zum Schulbesuch mit der Bildungsstatistik verglichen. Hierbei findet sich im Mikrozensus eine erhebliche Übererfassung der Schüler der gymnasialen Oberstufe. Sie hängt damit zusammen, dass Schüler beruflicher Schulen, an denen die Fachhochschulreife oder das Abitur erreicht werden können, entgegen den Definitionen des Mikrozensus als Schüler der gymnasialen Oberstufe erfasst sind. Der im Mikrozensuspanel niedrige Anteil von Absolventen mit Abitur oder Fachhochschulreife unter den Schülern der gymnasialen Oberstufe ist auf dieses Erfassungsproblem zurückzuführen.

In Bezug auf den Zusammenhang zwischen dem erreichten Bildungsstatus und den soziodemografischen sowie sozioökonomischen Variablen können aus diesen Gründen keine Abschlussquoten des Gymnasialschulbesuchs ermittelt werden. Unabhängig davon hängt der erreichte Status im Wesentlichen von der Alterskohorte ab. Der enge Zusammenhang zwischen (Abschluss-) Status und Alter spiegelt auch wider, dass jüngere Schüler häufig die gymnasiale Oberstufe verlassen.

Hinsichtlich der Panelausfälle weist die Variable Stellung im Beruf des Familienvorstands den engsten Zusammenhang mit dem Ausfall aufgrund räumlicher Mobilität auf. Dies spricht für die Annahme eines bedingt zufälligen Ausfalls („missing at random“, MAR) in Abhängigkeit von der beruflichen Stellung des Familienvorstands. Der zur Validierung der Modellbefunde vorgenommene Vergleich der Absolventenzahl allgemein bildender Schulen mit Allgemeiner Hoch-



schulreife des Mikrozensuspanels mit Ergebnissen der Bildungsstatistik zeigt, dass mithilfe der Ausfallgewichtung eine deutlich verbesserte Anpassung der Ergebnisse des Mikrozensuspanels erreicht werden kann.

Zusammenfassend erscheinen somit die durch Panelausfälle bei Verlaufsanalysen entstehenden Selektivitätsprobleme mithilfe von Gewichtungsfaktoren zwar nicht vollständig, aber doch weitgehend korrigierbar. Dies gilt allerdings nur, wenn die im Vergleich dazu gravierenderen Abgrenzungsprobleme zum Besuch und Abschluss der gymnasialen Oberstufe ignoriert werden.

So ernüchternd die Einbußen der Qualität der Angaben zum Schulbesuch für Bildungsforscher auch sein mögen, ist es gleichwohl für jede sachgerechte Auswertung wichtig zu wissen, in welcher Weise die Daten systematische Fehler aufweisen. Dies betrifft nicht nur die Paneldaten, sondern auch Querschnittsauswertungen. Auf Grundlage der hier berichteten Befunde können erste und einfache Konsequenzen gezogen werden. Beispielsweise kann nun bei Analysen zum Besuch der gymnasialen Oberstufe darauf verwiesen werden, dass diese Kategorie nicht nur Schüler allgemein bildender Schulen umfasst. Wenn zur gymnasialen Oberstufe auch berufliche Gymnasien gezählt werden, lässt sich mit der Beschränkung auf Schüler ab 18 Jahren außerdem eine gute Anpassung der Mikrozensusergebnisse an die Bildungsstatistik erreichen. Zugleich bieten sich den statistischen Ämtern mit diesen Ergebnissen Ansatzpunkte für die Verbesserungen der Fragebögen, der Intervieweranweisungen und der Interviewerschulung. In Bezug auf diese Aspekte ist zusammenfassend festzuhalten, dass mit dem Mikrozensuspanel ein neues Instrument zur Überprüfung der Datenqualität und der Qualitätssicherung vorliegt, das eine Vielzahl weiterer methodischer Auswertungsmöglichkeiten bietet.

Ab 2005 wird der Mikrozensus als kontinuierlich über das Jahr verteilte Befragung durchgeführt, sodass durch unterjährige Ergebnisse (Interviewzeitpunkte bzw. Berichtswochen, Monats- und Quartalsauswertungen) bessere Vergleichsmöglichkeiten zur Bildungsstatistik gegeben sind. Des Weiteren wird ab 2005 nach dem Schulbesuch in den letzten vier Wochen sowie in den letzten zwölf Monaten vor der jeweiligen Befragung gefragt. Zwar wird nach wie vor zwischen dem Besuch allgemein bildender vs. beruflicher Schulen unterschieden, jedoch wird ab 2005 zum Besuch einer beruflichen Schule explizit auch der jeweils erreichbare Abschluss genannt.<sup>51</sup> Eine weitere Aufklärung der hier zur amtlichen Bildungsstatistik festgestellten Verteilungsunterschiede muss anderen Analysen vorbehalten bleiben. Am besten geeignet für die Untersuchung von Klassifikationsfehlern sind sicherlich sorgfältige Pretests, da deren Ergebnisse helfen, solche Fehler zu vermeiden. Für die statistische Analyse können auch Wiederholungsbefragungen mit differenzierten Fragen zum Schulbesuch und Abschluss durchgeführt werden, bei denen hochqualifizierte Interviewer eingesetzt und Proxy-Interviews vermieden werden. Die daraus ermittelten Angaben lassen sich in Messfehlermodellen als eine Annäherung an „wahre Werte“ verwenden (vgl. hierzu Bassi et al. 2000; Biemer 2004; Särndal et al. 1997).

---

51 Erfreulicherweise wurden inzwischen die Fragen zum Schulbesuch geändert. Die hier diskutierten Klassifikationsfehler dürften im Mikrozensus ab 2008 nicht mehr auftreten, da zum Schulbesuch wieder die einzelnen Schularten erfragt werden. Dabei werden berufliche Schulen, die zur Fachhochschul- bzw. Hochschulreife führen, extra genannt. Anschließend werden die Klassenstufen erfragt.





## 6 Berufliche Ausbildungsverläufe bis zum Übergang ins Erwerbsleben

In diesem Kapitel liegt die Schwerpunktsetzung auf Bildungs- und Erwerbszusammenhängen, die einen wichtigen Kernbereich des Mikrozensus abdecken. Es soll geprüft werden, ob und in welcher Weise das Mikrozensuspanel dazu beitragen kann, den in diesem Themengebiet dringenden Bedarf an längsschnittlichen Informationen (Weißhuhn 2001: 57f., 80f.) zu befriedigen. Vor dem Hintergrund der großen Bedeutung eines beruflichen Ausbildungsabschlusses in Bezug auf spätere Lebens- und Einkommenslagen sind u. a. Fragen der hohen Quote von Ausbildungsabrechern angesprochen. Wie in der Einleitung erwähnt, ist auch zu klären, ob das Mikrozensuspanel „dynamische Kennziffern“ (Weißhuhn 2001: 65) zur Verwertbarkeit von Bildung auf dem Arbeitsmarkt liefern kann, die unter anderem für die Analyse des Wandels der Berufs- und Wirtschaftsstruktur benötigt werden. Dabei kann davon ausgegangen werden, dass insbesondere Berufsanfänger nach dem Ende ihres Ausbildungsvertrages von den gegenwärtigen Engpässen auf dem Arbeitsmarkt betroffen sind.

In einem Überblick empirischer Befunde zu Übergängen von Auszubildenden des dualen Systems ins Erwerbssystem stellte Steinmann (2000: 102f.) fest, dass zu diesen Fragen zwar viele Studien vorliegen, diese aber aufgrund jeweils unterschiedlicher Stichproben und Operationalisierungen wenig vergleichbar sind (siehe hierzu auch Gangl und Müller 1999). Zudem sind bei den von der Forschung verwendeten Längsschnittdaten für aussagekräftige Analysen oft die Fallzahlen zu klein oder es handelt sich um retrospektiv erhobene Verlaufsangaben, bei denen immer auch mit Erinnerungsfehlern gerechnet werden muss (Euwals und Winkelmann 2001; Franz et al. 2000). Für einen Test der Analysemöglichkeiten des Mikrozensus spricht allgemein neben der Stichprobengröße auch die infolge der Auskunftspflicht sehr hohe Ausschöpfung. Während mit den Beschäftigtenstichproben des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) umfangreiche Ereignisdaten verfügbar sind, die aber keine Angaben zum Haushaltskontext enthalten, können mit den als Haushaltsstichprobe durchgeführten Mikrozensen vielfältige Informationen zum Familien- und Haushaltshintergrund (z. B. Bildung und Beruf der Eltern) genutzt werden.

Ob diese allgemeinen Forschungspotenziale des Mikrozensus auch bei den zu einem Panel zusammengeführten Querschnittserhebungen, bei dieser Themenstellung und der dabei verwendeten kleinen und überdurchschnittlich räumlich mobilen Teilstichprobe der Auszubildenden gelten, hängt entscheidend davon ab, in welchem Maße durch räumliche Mobilität bedingte Verzerrungen vorliegen. Zur Validierung werden externe Daten der IAB-Beschäftigtenstichprobe herangezogen, die als zweiprozentige Stichprobe die Verläufe aller sozialversicherungspflichtig beschäftigten Arbeitnehmer, einschließlich räumlich mobiler Personen, enthalten.

Die Ergebnisse des vierten Kapitels zeigen, dass bei Übergang ins Erwerbssystem nach dem Abschluss der Berufsausbildung Jugendliche häufig das Elternhaus verlassen. Im Unterschied zu Besuchern allgemein bildender Schulen wurden für Besucher einer beruflichen Schule höhere Auszugswahrscheinlichkeiten geschätzt. Hinsichtlich der Befunde, dass das Ende einer Ausbildungsperiode sowie Betriebswechsel zum Anstieg der Auszugswahrscheinlichkeiten führen, ist zumindest teilweise zu erwarten, dass Ausfälle und Statuswechsel miteinander verbunden sind. Falls selektive Ausfälle vorliegen, ist zu prüfen, ob die daraus folgenden Verzerrungen mit Gewichtung- und Hochrechnungsfaktoren wirksam korrigiert werden können.

Das Kapitel ist wie folgt gegliedert: Der nächste Abschnitt befasst sich mit der Konstruktion vergleichbarer Abgrenzungen der Auszubildenden und Ausbildungsabschlüsse für das Mikrozensuspanel und die Beschäftigtenstichprobe. Mit Konzentration auf den Übergang 1996/97 folgen im zweiten Abschnitt zunächst deskriptive Vergleiche der Verteilungen der Auszubildenden, der Ausbildungsabschlüsse und –abbrüche im Mikrozensuspanel und in der Beschäftigtenstichprobe. In diesem Abschnitt werden Hypothesen zum Ausfall bzw. zur räumlichen Mobilität des vierten Kapitels aufgegriffen und für die hier untersuchte Population der Auszubildenden des dualen Systems konkretisiert. Darauf aufbauend folgen statistische Analysen zum Ausfall. Tests zur Einsatzmöglichkeit von Gewichtungsfaktoren sind Gegenstand des dritten Abschnitts. Im fünften Abschnitt finden sich deskriptive Analysen zu den Chancen der Beschäftigung im erlernten Beruf. Das Kapitel schließt mit einer Zusammenfassung der Hauptergebnisse.

## **6.1 Abgrenzung der Auszubildenden und der Ausbildungsabschlüsse**

Unter den beruflichen Ausbildungen im beruflichen Bildungssystem der Sekundarstufe II sowie in Fachhochschulen und Hochschulen nimmt die berufliche Ausbildung im dualen System quantitativ eine zentrale Stelle ein. Nach Ergebnissen des Mikrozensus 1996 absolvieren beispielsweise rund ein Drittel der 17- bis 19-jährigen bzw. 15 Prozent der 15- bis 24-jährigen Jugendlichen eine Ausbildung im dualen System. Seit dem Berufsbildungsgesetz von 1969 sind die Ausbildungsinhalte für derzeit rund 350 anerkannte Ausbildungsberufe festgelegt. In der Ausbildung wird die theoretische Wissensvermittlung in Berufsschulen, die im Teilzeitunterricht besucht werden, mit der praktischen Ausbildung im Betrieb kombiniert. Zwischen dem Auszubildenden (früher „Lehrling“) und dem Ausbildungsbetrieb wird ein Ausbildungsvertrag abgeschlossen. Die Ausbildungsdauer beträgt meistens drei Jahre, sie variiert aber nach den Ausbildungsberufen. Für Auszubildende mit Abitur oder einer beruflichen Vorbildung (z. B. Berufsfachschule) kann die Ausbildungsdauer um ein Jahr verkürzt werden.

Nicht zu den nach dem Berufsbildungsgesetz definierten Auszubildenden im dualen System (Legaldefinition) zählen Praktikanten, Volontäre, Beamte im Vorbereitungsdienst und Umschüler. Gleichfalls zählen Ausbildungen an beruflichen Schulen, wie beispielsweise zum Krankenpfleger, nicht dazu, obwohl sie in Bezug auf die Organisation theoretischer und praktischer Curricula Ähnlichkeiten mit der Ausbildung im dualen System besitzen. Ebenfalls zählen Ausbildungen in überbetrieblichen Ausbildungswerkstätten nicht zum dualen System.

Die der obigen Legaldefinition bzw. dem Berufsbildungsgesetz entsprechenden Auszubildenden des dualen Systems (siehe BMBF 2002: 114; Statistisches Bundesamt 2000: 10-14) werden von den zuständigen Stellen (i. d. R. Kammern) erfasst. Die in der Berufsbildungsstatistik für die Population berichteten Ergebnisse der jeweils zum 31.12. eines Jahres bestehenden Ausbildungsverhältnisse und Abschlüsse etc. kämen zwar in erster Linie als Referenzdaten in Frage, sie enthalten jedoch auch eine unbekannte Zahl von Auszubildenden, die in außerbetrieblichen Ausbildungsstätten ausgebildet werden (Beicht et al. 2003: 4). Dazu zählen insbesondere Maßnahmen im Rahmen von Sonderprogrammen des Bundes und der Länder, der Bundesagentur für Arbeit oder anderer Träger. Nach Schätzungen sind neun Prozent der bis Ende September 2004 neu abgeschlossenen Ausbildungsverträge außerbetrieblich (BMBF 2005: 632).<sup>52</sup>

52 Der Anteil außerbetrieblicher Ausbildungsverträge ist in den neuen Bundesländern (einschl. Berlin) mit 28 Prozent wesentlich höher als in den alten Bundesländern mit rund vier Prozent.

### 6.1.1 Ausgangsinformationen im Mikrozensuspanel

Im Mikrozensus können Auszubildende mithilfe der Angaben zur Stellung im Beruf und zum Schulbesuch ermittelt werden. Ergänzend liegen für die Befragten der in den Mikrozensus integrierten EU-Arbeitskräfteerhebung (0,45%-Substichprobe) Angaben zur Aus- und Weiterbildung vor. Der zeitliche Vergleich der Angaben zum beruflichen Ausbildungsabschluss zeigt, ob die Ausbildung erfolgreich abgeschlossen wurde.

Bei der Frage zur Stellung im Beruf geben Befragte an, ob sie kaufmännisch, technisch oder gewerblich Auszubildende sind. Den gewerblich Auszubildenden sind handwerklich und landwirtschaftlich Auszubildende zugeordnet. Im Unterschied zur obigen Legaldefinition zählen gemäß des Interviewer-Handbuchs im Mikrozensus auch Praktikanten, Volontäre und Schüler an Schulen des Gesundheitswesens, die gleichzeitig praktisch ausgebildet werden, zu den Auszubildenden. Zur Abgrenzung der Auszubildenden des dualen Systems können die Angaben zum Berufsschulbesuch und zum Beruf herangezogen werden.

Bei dieser Abgrenzung von Auszubildenden des dualen Systems ist jedoch in Bezug auf den Schulbesuch zu beachten, dass die Angabe „Berufliche Schule“ neben der Berufsschule auch weitere berufliche Schulen umfasst (u. a. Berufsgrundbildungsjahr, Berufsaufbauschulen, berufliche Gymnasien, Berufsakademien, Schulen des Gesundheitswesens).

Die Berufsangaben sind nach der Klassifizierung der Berufe 1992 vercodet und liegen als Berufsordnungen (Dreisteller) vor. Für eine korrekte Abgrenzung „anerkannter Ausbildungsberufe“ werden jedoch die differenzierteren Berufsklassen (Viersteller) benötigt. Auf Basis der Berufsordnungen ist somit die Erfassung anerkannter Ausbildungsberufe nur näherungsweise möglich.

In Bezug auf die Ermittlung des erfolgreichen Ausbildungsabschlusses ist zu beachten, dass Absolventen einer Ausbildung im dualen System zusammen mit Absolventen von Berufsfachschulen (u. a. Krankenpflegeschulen) in der gemeinsamen Kategorie „Abschluss einer Lehrausbildung oder gleichwertiger Berufsfachschulabschluss“ ausgewiesen sind.<sup>53</sup>

Die Fragen zur Aus- und Weiterbildung (gegenwärtig bzw. ein Jahr zuvor) beinhalten u. a. Informationen darüber, ob es sich um eine Lehrausbildung, ein berufliches Praktikum, eine Fortbildung oder Umschulung handelt. Diese Informationen liegen zwar nur für die Substichprobe der EU-Arbeitskräfteerhebung vor, sie können aber dazu verwendet werden, die Abgrenzung von Auszubildenden des dualen Systems zu überprüfen.

Da sich die Frage nach der beruflichen Stellung nur an Erwerbstätige richtet, die außerbetrieblichen Auszubildenden als Teilnehmer der oben genannten Sonderprogramme jedoch nicht erwerbstätig sind, können außerbetriebliche Auszubildende des dualen Systems im Mikrozensus nicht von Besuchern anderer beruflicher Schulen unterschieden werden. Die Berufsbildungsstatistik, die außerbetriebliche Auszubildende enthält, ist damit für Vergleiche des Mikrozensuspanels wenig geeignet. Hinzu kommt der zeitliche Abstand zwischen den Ergebnissen der Berufsbildungsstatistik zum 31.12. des Vorjahres und der Mikrozensusbefragung im April eines Jahres. Vor dem Hintergrund, dass etwa ein Viertel der Ausbildungsverträge vorzeitig, zumeist innerhalb der Probezeit, aufgelöst werden (Statistisches Bundesamt 2000: 9), ist dieser zeitliche Abstand der Berichtszeiträume beider Datenquellen zu groß. Aus diesen Gründen kann beim Vergleich der Ergebnisse des Mikrozensuspanels mit der Berufsbildungsstatistik zwischen Ausfall bzw. Wegzug aus dem Auswahlbezirk und Ausbildungsabbruch nicht unterschieden werden. Als Referenzstatistik wird deshalb die Beschäftigtenstichprobe verwendet.

53 Das Berufsvorbereitungsjahr und der Abschluss der Berufsfachschule werden erst ab dem Mikrozensus 1999 extra ausgewiesen.

### 6.1.2 Ausgangsinformationen in der Beschäftigtenstichprobe

In der Beschäftigtenstichprobe können zur Abgrenzung von Auszubildenden die Merkmale Stellung im Beruf, Beruf und Bildungsqualifikation genutzt werden. Im Unterschied zum Mikrozensus stammen die Angaben nicht von den Beschäftigten (bzw. bei Proxy-Interviews von anderen Personen im Haushalt), sondern von den Arbeitgebern. Die prozessproduzierten Daten erlauben die Ermittlung des Auszubildenden- bzw. Erwerbsstatus entsprechend den Berichtswochen des Mikrozensuspanels. Allerdings entstehen Ausfälle, falls innerhalb der Panelzeiträume (Berichtswochen) keine sozialversicherungspflichtige Beschäftigung vorliegt.

Ähnlich wie im Mikrozensus umfasst die Kategorie „in Ausbildung“ der Variablen Stellung im Beruf auch Praktikanten (einschl. Anlernausbildungen), Volontäre und Umschüler sowie Ausbildungen im Gesundheitswesen. Erst ab April 1999, nach Änderung des Meldeverfahrens, sind Praktikanten und Werkstudenten extra ausgewiesen. Mithilfe der Variablen Personengruppe sind Auszubildende und Praktikanten usw. aber unterscheidbar. Die Angabe „Auszubildende“ umfasst ausschließlich betriebliche Ausbildungen, sodass vergleichbar zum Mikrozensus außerbetriebliche Ausbildungen ausgeschlossen werden können.<sup>54</sup>

In der Beschäftigtenstichprobe sind die Berufsangaben nach der Klassifizierung der Berufe 1998 der Bundesagentur für Arbeit vercodet, die sich von der Berufsklassifikation des Statistischen Bundesamtes unterscheidet. Zudem sind die Berufskategorien aus Datenschutzgründen in 130 Kategorien stärker zusammengefasst als im Mikrozensus. Sie liegen teilweise auf der Ebene von Berufsordnungen (Dreisteller), teilweise aber nur als zweistellige Berufsgruppen vor.

Seit 1992 müssen die Betriebe das Ende einer Ausbildung melden (Schwerdt und Bender 2003: 48). Ein Codewechsel in der Variablen Stellung im Beruf indiziert somit das Ausbildungsende. Ob ein erfolgreicher Ausbildungsabschluss vorliegt, kann durch den Vergleich der Angaben zur Schul-/Berufsausbildung während und nach der Ausbildungsphase festgestellt werden. In dieser Variablen sind allgemein bildende und berufliche Abschlüsse kombiniert. Vergleichbar zum Mikrozensus umfasst die Kategorie „abgeschlossene Berufsausbildung“ (Lehr-) Ausbildungen in anerkannten Ausbildungsberufen, aber auch Abschlüsse einer Berufsfachschule oder Fachschule.

### 6.1.3 Konstruktion vergleichbarer Abgrenzungen

Die Beschäftigtenstatistik bietet eine nicht durch Wegzüge verzerrte Analyse der Übergänge zwischen Ausbildung und erstem Beruf und ermöglicht somit die Untersuchung des Selektionsbias im Mikrozensuspanel. Damit aber Selektivitätseffekte nicht durch andere Differenzen zwischen den Datenquellen überlagert werden, sollten die jeweils verwendeten Variablen möglichst vollständig vergleichbar sein.

Für die Untersuchung der Chancen einer Beschäftigung im erlernten Beruf nach dem Ende der Ausbildung ist im ersten Schritt die Risikomenge der Auszubildenden im dualen System abzugrenzen. Im zweiten Schritt kann dann der erfolgreiche Ausbildungsabschluss ermittelt werden. Die zur Abgrenzung der Auszubildenden und der Ausbildungsabschlüsse verwendeten Merkmale sind in Abbildung 6.1 zusammengestellt.

<sup>54</sup> Da in der Beschäftigtenstatistik nur außerbetriebliche Ausbildungen enthalten sind, die nach dem Sozialgesetzbuch gefördert werden, ist im Vergleich zur Berufsbildungsstatistik mit einer Untererfassung außerbetrieblicher Ausbildungen zu rechnen.

Abgrenzung	Mikrozensuspanel	Beschäftigtenstichprobe
Auszubildende des dualen Systems	<p>Stellung im Beruf = Kaufm. / Techn. / Gewerbl. Auszubildende  <i>Inkl. Praktikanten, Volontäre, Schüler an Schulen des Gesundheitswesens. Nur Erwerbstätige (betriebliche Ausbildungen).</i></p> <p>Schulbesuch = Berufliche Schule  <i>Berufsschulen, Berufsgrundbildungsjahr, Berufsvorbereitungsjahr, Berufsaufbauschulen, Fachoberschulen, berufl. Gymnasien, Kollegschen (NRW), Berufsfachschulen, Fachschulen, Fach-/Berufsakademien, Schulen des Gesundheitswesens</i></p> <p>Beruf = anerk. Ausbildungsberuf  <i>Recodierung von Berufsordnungen (KldB 1992 StBA) in Berufsabschnitte (KldB 1988 BA)</i></p> <p>Alter (Geburtsjahr) <math>\leq 24</math> Jahre  <i>[Ausschluss potenzieller Fortbildungen und Umschulungen]</i></p>	<p>Stellung im Beruf = In Ausbildung  <i>Inkl. Anlernlinge, Praktikanten, Volontäre, Ausbildungen im Gesundheitswesen, Umschüler.</i></p> <p>Personengruppe = Auszubildende  <i>Nur betriebliche Ausbildungen (Merkmal nur in Beschäftigtenhistorik. Ab 1999 in Westdeutschland: ohne Praktikanten und Werkstudenten).</i></p> <p>Beruf = anerk. Ausbildungsberuf  <i>Recodierung von Berufsabschnitten und Berufsgruppen (KldB 1988 BA) in Berufsabschnitte</i></p> <p>Alter (Geburtsjahr) <math>\leq 24</math> Jahre</p>
<p>Beruflicher Ausbildungsabschluss  <i>[Nur für Erstausbildungen ermittelbar, d. h. wenn ein Jahr zuvor eine berufliche Schule besucht wurde und kein beruflicher Abschluss vorlag.]</i></p>	<p>Vergleich zur Erhebung ein Jahr zuvor:</p> <p>Stellung im Beruf <math>\neq</math> Auszubildender</p> <p>Schulbesuch <math>\neq</math> Berufliche Schule</p> <p>Beruflicher Abschluss = Lehrausbildung  <i>Lehrausbildung oder gleichwertiger Berufsfachschulabschluss. Als Lehrausbildung gilt eine Ausbildung von mind. zwei Jahren.</i></p>	<p>Vergleich zur Meldung ein Jahr zuvor:</p> <p>Stellung im Beruf <math>\neq</math> in Ausbildung          Personengruppe <math>\neq</math> Auszubildende</p> <p>Schul-/Berufsausbildung = (...) mit Berufsausbildung  <i>Ausbildungen in anerkannten Lehrberufen i. S. des Berufsbildungsgesetzes und Abschlüsse von Berufsfach-/Fachschulen.</i></p>

Abbildung 6.1: Kurzbeschreibung zur Abgrenzung der Auszubildenden im dualen System und der Ausbildungsabschlüsse

Im Mikrozensus und in der Beschäftigtenstichprobe sind Auszubildende in der Variablen Stellung im Beruf weitgehend gleich erfasst. Um in der Beschäftigtenstatistik außerbetriebliche Ausbildungen auszuschließen, ist zusätzlich das Merkmal Personengruppe zu verwenden, das vor der regulären Aufnahme als Meldemerkmal bis 1999 aus bereits vorhandenen Informationen durch das IAB gebildet wurde. Aufgrund der hohen Übereinstimmung der Variablen Stellung im Beruf in beiden Daten stellt sich in Bezug auf die Abgrenzung von Auszubildenden des dualen Systems beim Mikrozensus vorwiegend die Frage, wie Werkstudenten und Praktikanten usw. sowie Aus-

bildungen außerhalb des dualen Systems (z. B. Krankenpflegeausbildung) ausgeschlossen werden können. Testauswertungen der Angaben zur Aus- und Weiterbildung der Substichprobe (EU-Arbeitskräftestichprobe) des Jahres 1996 zeigten, dass bei einer Beschränkung auf die bis 24-Jährigen eine Konzentration auf berufliche Erstausbildungen möglich ist und Umschulungen und Fortbildungen weitestgehend ausgeschlossen werden können (Schimpl-Neimanns 2006b: 12).

Die genaue Abgrenzung der Auszubildenden des dualen Systems ist nur mithilfe der vierstelligen Berufsklassen der Berufsklassifikationen der Bundesagentur für Arbeit bzw. des Statistischen Bundesamtes möglich. Die beiden Klassifikationssysteme sind zudem nur auf der Ebene der Viersteller vergleichbar. Die zur Verfügung stehenden Daten enthalten jedoch maximal Berufsordnungen (Dreisteller) und sind im Fall der Beschäftigtenstichprobe aus Datenschutzgründen teilweise weiter zu Berufsabschnitten zusammengefasst. Eine Berufskategorie des Mikrozensuspanels oder der Beschäftigtenstichprobe kann näherungsweise der Klasse anerkannter Ausbildungsberufe zugeordnet werden, falls sie vierstellige Berufsklassen mit anerkannten Ausbildungsberufen enthält. Dabei entstehen jedoch Klassifikationsfehler, da diese Zuordnung auch Berufe einschließt, die nicht zu den anerkannten Ausbildungsberufen gehören.

Aufgrund der im Vergleich zum Mikrozensuspanel (342 Kategorien) größeren Berufsangaben in der Beschäftigtenstichprobe (130 Kategorien) ist bei der Beschäftigtenstichprobe mit einem höheren Anteil von Fehlklassifikation zu rechnen. Deshalb werden im Folgenden die differenzierteren Berufscodes des Mikrozensus zunächst zu größeren Berufscodes der Beschäftigtenstichprobe zusammengefasst.<sup>55</sup> Die anschließende Zuweisung zur Klasse anerkannter Ausbildungsberufe ist dann für beide Datenbasen einheitlich. Testauswertungen haben gezeigt, dass sich bei diesem Vorgehen im Vergleich zu der für jede Datei bzw. für jede Berufsklassifikation getrennten Recodierung deutlich ähnlichere Verteilungen der Berufe von Auszubildenden im Mikrozensuspanel und in der Beschäftigtenstatistik ergeben.

Das Abschlussjahr des höchsten beruflichen Abschlusses wird im Mikrozensus nur im Rahmen der Unterstichprobe im vierjährigen Turnus ohne Auskunftspflicht erfragt. Im Mikrozensuspanel sind jedoch nur die 1996 erhobenen Angaben enthalten. Die Meldungen in der Beschäftigtenstatistik beziehen sich nur auf die Art des Schulabschlusses. Für beide Datenquellen können das Ende der Ausbildung und der erfolgreiche Abschluss aus diesen Gründen nur im Vergleich der Querschnittsangaben ermittelt werden. Einschränkungen ergeben sich daraus, dass sich die Angaben auf den höchsten erreichten Abschluss beziehen und die Abschlusskategorie neben der Lehre auch gleichwertige Berufsfachschulabschlüsse enthält. Für Berufsfachschüler sowie Personen mit einem Fachschulabschluss oder höheren Abschluss, die zusätzlich eine Ausbildung im dualen System absolviert haben, kann der Lehrabschluss durch den Vergleich der Querschnittsangaben nicht festgestellt werden. Der berufliche Ausbildungsabschluss ist folglich nur für Personen in einer Erstausbildung ermittelbar, d. h. wenn kein anderer beruflicher Ausbildungsabschluss vorliegt.

Um die Güte der so abgegrenzten Auszubildenden des dualen Systems und der Ausbildungsabschlüsse zu prüfen, enthält Tabelle 6.1 neben den Verteilungen der Berufe im Mikrozensuspanel und der Beschäftigtenstichprobe weitere Merkmale der 15- bis 24-jährigen Auszubildenden in anerkannten Ausbildungsberufen. Alle Regionalangaben der Beschäftigtenstichprobe beziehen sich auf den Betriebsort. Das Bundesland der Arbeitsstätte lässt sich im Mikrozensus ebenfalls ermitteln, ab 1997 jedoch nur für die Unterstichprobe.

55 Die Recodierung erfolgte auf Basis eines Umsteigeschlüssels für die Berufsklassen (Viersteller) der Berufsklassifikation des Statistischen Bundesamtes und der Bundesagentur für Arbeit. Für dessen Übermittlung danke ich Hans Dietrich.



Tabelle 6.1: Merkmalsverteilungen in der Beschäftigtenstichprobe und im Mikrozensuspanel; 15- bis 24-jährige Auszubildende in anerkannten Ausbildungsberufen im April 1996 – in Prozent

Merkmal	Beschäftigten- stichprobe	Mikro- zensuspanel	Differenz
Ort der Arbeitsstätte <sup>1)</sup>			
West	77,0	75,6	1,4
Ost (inkl. West-Berlin)	23,0	24,4	-1,4
Geschlecht			
männlich	57,3	58,7	-1,4
weiblich	42,7	41,3	1,4
Alter 1996 (Geburtsjahr)			
≤16	1,7	1,7	0,0
17	9,8	11,6	-1,8
18	20,4	22,1	-1,7
19	23,5	23,9	-0,4
20	18,9	17,3	1,6
21	11,6	11,6	0,0
22	7,4	5,9	1,5
23-24	6,7	5,8	0,9
Staatsangehörigkeit (nur West) <sup>2)</sup>			
deutsch (inkl. Ausländer in Ostdeutschland)	92,1	88,3	3,8
nicht deutsch	7,9	11,7	-3,8
Allgemeine und berufliche Ausbildung			
Haupt-/Realschule (inkl. POS) ohne Berufsausbildung	83,6	79,1	4,5
Haupt-/Realschule (inkl. POS) mit Berufsausbildung	4,1	4,1	0,0
Abitur ohne Berufsausbildung	8,1	11,8	-3,7
Abitur mit Berufsausbildung	0,5	0,8	-0,3
Fachhochschul-/Hochschulabschluss	0,4	0,3	0,1
Ausbildung unbekannt	3,4	3,9	-0,5
Beruf (Berufsabschnitte)			
Pflanzenbauer, Tierzüchter, Fischereiberufe	2,5	2,7	-0,2
Bergleute, Mineralgewinner	0,2	0,3	-0,1
Keramiker, Glasmacher, Chemiearb., Kunststoffverarb.	0,7	0,7	0,0
Papierhersteller, -verarbeiter, Drucker	0,7	0,6	0,1
Holzaufbereiter, Metallherzeuger, -bearbeiter	0,9	0,9	0,0
Schlosser, Mechaniker	18,0	18,2	-0,2
Elektriker, Montierer und Metallberufe	7,1	6,1	1,0
Textil- u. Bekleidungsberufe, Lederherst., -verarbeiter	0,5	0,5	0,0
Ernährungsberufe	3,9	3,6	0,3
Bauberufe	6,9	7,6	-0,7
Bau-, Raumausstatter, Polsterer	1,8	2,0	-0,2
Tischler, Modellbauer	3,3	3,3	0,0
Maler, Lackierer	2,9	2,7	0,2



Merkmal	Beschäftigten- stichprobe	Mikro- zensuspanel	Differenz
Warenprüfer, Versandfertigm., Hilfsarb., Maschinisten	0,1	0,1	0,0
Ingenieure, ..., Techniker	2,1	1,7	0,4
Warenkaufleute	11,0	11,3	-0,3
Dienstleistungskaufleute	6,0	5,8	0,2
Verkehrsberufe	0,8	0,9	-0,1
Organisations-, Verwaltungs-, Büroberufe	17,1	18,0	-0,9
Schriftwerkschaffende	0,5	0,4	0,1
Gesundheitsberufe	6,6	6,1	0,5
Allgemeine Dienstleistungsberufe	6,5	6,2	0,3
Wirtschaftszweig <sup>3)</sup>			
Landwirtsch., Gartenbau, Gärtnerei, Energie u. Bergbau	2,4	3,6	-1,2
Grundstoff-, Güterproduktion	3,8	3,1	0,7
Investitionsgüterproduktion	14,3	14,8	-0,5
Verbrauchsgütergewerbe	5,4	3,8	1,6
Nahrungs- und Genussmittelgewerbe	3,6	2,0	1,6
Bauhauptgewerbe	7,6	7,3	0,3
Ausbaugewerbe	10,1	12,2	-2,1
Distributive Dienstleistungen	14,3	16,5	-2,2
Verkehr, Nachrichtenübermittlung	3,5	3,3	0,2
Vorwiegend wirtschaftsbezogene Dienstleistungen	11,2	10,8	0,4
Vorwiegend haushaltsbezogene Dienstleistungen	7,3	9,1	-1,8
Gesellschaftsbezogene Dienstleistungen	13,0	8,6	4,4
Gebietskörperschaften, Sozialversicherung	3,5	4,9	-1,4
Insgesamt (hochgerechnete Fallzahlen in 1.000 = 100 %)	1.213,2	1.041,7	-14,1 %

Quelle: Mikrozensuspanel 1996-1999 (Bevölkerung am Hauptwohnsitz; gebundene Hochrechnung), Beschäftigtenstichprobe (IABS-R01); eigene Berechnungen.

<sup>1)</sup> Mikrozensuspanel: ohne 6,1 % fehlende Angaben zum Ort der Arbeitsstätte

<sup>2)</sup> Beschäftigtenstichprobe: ohne 0,4 % fehlende Angaben zur Staatsangehörigkeit

<sup>3)</sup> Beschäftigtenstichprobe: ohne 0,6 % fehlende Angaben zum Wirtschaftszweig

Gemäß den so vorgenommenen Abgrenzungen absolvierten im April 1996 nach dem Mikrozensuspanel hochgerechnet 1.041.700 Personen eine Ausbildung im dualen System. Auf Basis der Beschäftigtenstichprobe werden 1.213.200 Auszubildende geschätzt. Gegenüber der Beschäftigtenstichprobe wird mit dem Mikrozensuspanel damit die Zahl der Auszubildenden um 14 Prozent unterschätzt. In dieser Größenordnung liegt auch die Untererfassung der Schüler beruflicher Schulen des Mikrozensuspanels im Vergleich mit der Berufsschulstatistik (Schimpl-Neimanns 2005: 29). Die Ursache dafür könnte in der Frageformulierung, der Fokussierung auf den Interviewzeitpunkt und der gebundenen Hochrechnung des Mikrozensus liegen, die (bis 2004) nicht nach Alter differenziert (siehe dazu auch Abschnitt 5.3.1, Seite 84). Eine Untererfassung von Auszubildenden kann darin begründet sein, dass diese Gruppe in der Berichtswoche oder während des Interviews keine Berufsschule besuchte, sich aber zu diesen Zeitpunkten noch in einer dualen Ausbildung befand. Eine Untererfassung würde auch dann auftreten, falls zwischen der

Berichtswoche und dem Interview die Ausbildung beendet wurde und die Befragten ihre Antwort auf den Interviewzeitpunkt beziehen. Ohne Kenntnis des Interviewzeitpunkts reichen jedoch die in den Daten vorhandenen Informationen für eine Überprüfung solcher Vermutungen nicht aus.

Abgesehen von der Gesamtabweichung sind die Verteilungsdifferenzen zwischen beiden Datenquellen insbesondere bei der Arbeitsstätte und beim Geschlecht der Auszubildenden vernachlässigbar gering. Etwas stärkere Abweichungen von maximal rund vier Prozentpunkten zeigen sich bei der Staatsangehörigkeit,<sup>56</sup> beim Bildungsabschluss und beim Wirtschaftszweig. Haupt- und Realschüler ohne Berufsausbildung sind im Mikrozensus im Vergleich zur Beschäftigtenstichprobe unterrepräsentiert (79,1 % vs. 83,6 %), Abiturienten ohne Berufsausbildung dagegen überrepräsentiert (11,8 % vs. 8,1 %). Hierbei ist aber zu beachten, dass die Angaben der Betriebe zum Ausbildungsabschluss in der Beschäftigtenstichprobe versicherungsrechtlich ohne Belang sind. In Bezug auf die Wirtschaftszweige liegt im Mikrozensuspanel eine Untererfassung von rund vier Prozent bei den gesellschaftsbezogenen Dienstleistungen (8,6 % vs. 13 %) vor. Allerdings ist die Vergleichbarkeit gravierend eingeschränkt. Insbesondere dürfte die Abweichung bei den gesellschaftsbezogenen Dienstleistungen darauf zurückzuführen sein, dass die im Mikrozensus verwendete Klassifikation die frühere Unterscheidung nach Sektoren nicht mehr zulässt.<sup>57</sup>

Wenngleich die Verteilungen in Tabelle 6.1 für eine gut vergleichbare Abgrenzung der Auszubildenden des dualen Systems im Mikrozensuspanel und in der Beschäftigtenstichprobe sprechen, ist auch darauf hinzuweisen, dass stärkere Abweichungen bei Berufen von Auszubildenden zu finden sind, die keinen anerkannten Ausbildungsberuf ausüben (hier ohne Nachweis; siehe Schimpl-Neimanns 2006b: 16). Insbesondere zeigen sich gravierende Unterschiede bei Auszubildenden mit noch nicht feststehendem Ausbildungsberuf. Diese Differenzen dürften mit systematischen Unterschieden in der Berufscodierung von Umfragedaten und prozessproduzierten Daten zusammenhängen. Neben den durch Proxy-Interviews beim Mikrozensus<sup>58</sup> entstehenden Problemen sind allgemein bei Umfragen die Schwierigkeiten von Befragten bei der Unterscheidung zwischen schulischen und betrieblichen Ausbildungen zu beachten.<sup>59</sup> Zusammenfassend kann aber festgehalten werden, dass die Vergleichbarkeitsprobleme zwar nicht ganz ausgeschaltet, jedoch erheblich reduziert werden können. Dies spricht dafür, dass die Risikopopulationen der Auszubildenden des Ausgangsjahres 1996 in beiden Daten weitgehend gleich erfasst sind.

56 In der Beschäftigtenstichprobe wird die Staatsangehörigkeit (Deutsche / Nicht Deutsche) nur für Westdeutschland ausgewiesen.

57 Im Mikrozensus sind die Angaben nach der Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 1993), des Statistischen Bundesamtes vercodet. Die Daten der Beschäftigtenstichprobe sind nach der Klassifikation der Bundesanstalt für Arbeit (Ausgabe 1973) verschlüsselt sowie aus Datenschutzgründen stark vergrößert. Beide Klassifikationen weichen systematisch voneinander ab. Der in Tabelle 6.1 berichteten Verteilung der Wirtschaftszweige des Mikrozensuspanels liegt eine grobe Umcodierung zur Klassifikation der Wirtschaftszweige (Ausgabe 1979) sowie eine anschließende Zusammenfassung zu ähnlichen Gruppen wie in der Beschäftigtenstichprobe zugrunde. Da allerdings in der WZ 1993 keine Sektoren (Unternehmen, Private Haushalte und Organisationen ohne Erwerbszweck, Staat) unterschieden werden können, ist die Vergleichbarkeit mit den Systematiken bis zur WZ 1979 nicht mehr gegeben.

58 Bei diesen Proxy-Interviews ist zu vermuten, dass die Auskunft gebenden Personen (z. B. Eltern) nicht immer über die für eine korrekte Berufsklassifikation benötigten Informationen verfügen.

59 Beispielsweise zeigte sich in der BIBB/IAB-Erhebung 1998/99, dass ein Viertel der Befragten, die sich selbst bei einer „betrieblichen Ausbildung“ eingestuft hatten, nach Kontrollen mithilfe von differenzierten Angaben zur Ausbildung als Absolventen einer schulischen Ausbildung einzustufen waren (Hall o. J.). Dies betrifft überwiegend die Berufe im Bereich der Krankenpflege und Erziehung.

Nachfolgende Übergänge bzw. Wechsel des Ausbildungsstatus und Fragen der Stichproben-selektivität sind damit untersuchbar.

## **6.2 Auszubildende und Ausbildungsabschlüsse im Mikrozensuspanel und in der Beschäftigtenstichprobe**

Zur Vergleichbarkeit folgen in diesem Abschnitt weitere deskriptive Untersuchungen sowie erste Analysen zu den Übergängen von der Ausbildung zum Ausbildungsabschluss. Daran anschließend werden die Fragen selektiver Ausfälle im Mikrozensuspanel bearbeitet.

In Tabelle 6.2 sind die Angaben zu den Auszubildenden 1996, die Übergänge und Abschlüsse bis 1997 sowie die Ausfälle bzw. Wegzüge und Zugänge zusammengestellt. Zwar sind die Unterfassungen im Mikrozensuspanel 1998 und 1999 geringer als 1996/97, da aber die Übergänge 1997/98 und 1998/99 ähnliche Verteilungen aufweisen (Schimpl-Neimanns 2006b: 64-65), kann sich die Gegenüberstellung der Ergebnisse des Mikrozensuspanels und der Beschäftigtenstichprobe im Folgenden auf den Übergang 1996/97 beschränken. Außer bei Ausfällen wird hierbei vereinfacht danach unterschieden, ob eine abgeschlossene Lehrausbildung vorliegt oder nicht. Bei den Abgängen kann durch den zeitlichen Vergleich des Bildungsstatus festgestellt werden, ob die Ausbildung erfolgreich abgeschlossen wurde, sofern nicht bereits 1996 ein Lehrabschluss oder ein gleichwertiger Berufsfachschulabschluss vorlag. Die Anteile von Auszubildenden mit bereits vorhandenem beruflichen Ausbildungsabschluss sind in beiden Datenquellen mit jeweils rund fünf Prozent gering.

Auf Basis der Beschäftigtenstichprobe können für die Berichtswoche im April 1997 zunächst nur jene Personen ausgewertet werden, für die zu diesem Termin eine Meldung zur Sozialversicherung vorliegt. Dies trifft für insgesamt rund acht Prozent der Auszubildenden im April 1996 nicht zu. Die Zahlen zum Abschluss bei den Abgängen enthalten in diesem Fall die der Berichtswoche April 1997 am nächsten liegenden Angaben. Für rund 850 Auszubildende konnte auf diese Weise als letzter Status nur ermittelt werden, dass sie vor der Meldelücke Auszubildende waren. Es erscheint aber plausibel, diese Gruppe zu den Ausbildungsabbrechern zu zählen.<sup>60</sup>

Zum Vergleich sind die Abgänge (ohne Wegzüge bzw. Ausfälle) des Mikrozensuspanels für die Personengruppe dargestellt, die in der Beschäftigtenstichprobe ohne Meldung bleiben; dies sind nicht sozialversicherungspflichtig Beschäftigte (Beamte, Selbstständige, mithelfende Familienangehörige, ausschließlich geringfügig Beschäftigte und Nichterwerbspersonen) sowie Wehr- und Zivildienstleistende. Diese Abgänge betragen im Mikrozensuspanel 6,8 Prozent gegenüber 7,9 Prozent in der Beschäftigtenstichprobe. Nimmt man beim Mikrozensuspanel an, dass sich die Anteile nicht sozialversicherungspflichtig Beschäftigter bei den ausgefallenen bzw. weggezogenen Personen nicht von den räumlich immobilen Personen unterscheiden, d. h. dass es sich in dieser Hinsicht um einen völlig zufälligen Ausfall handelt, wird der Anteil von Personen mit nicht sozialversicherungspflichtiger Beschäftigung, bezogen auf alle Auszubildenden 1996,

60 Die Zahl der Personen mit Meldelücken erhöht sich auf rund 900, wenn auch Auszubildende in einer beruflichen Zweitausbildung hinzu genommen werden. Für jene Personen, bei denen als Status vor der Meldelücke kein Abschluss ermittelt werden konnte, ergab eine Überprüfung, dass 86 Prozent der Personen mit nachfolgender sozialversicherungspflichtiger Beschäftigung auch zu einem späteren Zeitpunkt keinen Berufsausbildungsabschluss aufweisen. Es liegt deshalb nahe, die gesamte Gruppe zu den Ausbildungsabbrechern zu zählen. Die dabei nicht ausschließbare Fehlzuordnung reduziert den Anteil erfolgreicher Lehrabsolventen bzw. vergrößert den Anteil von Ausbildungsabbrechern aber nur um maximal 0,7 Prozent.

auf insgesamt 8,4 Prozent geschätzt. Die Differenz zwischen Beschäftigtenstichprobe und Mikrozensuspanel beträgt dann lediglich 0,5 Prozentpunkte.

Tabelle 6.2: Übergänge 1996/97 der 15- bis 24-jährigen Auszubildenden in anerkannten Ausbildungsberufen des Jahres 1996

Bestand und Bestandsentwicklung Beruflicher Abschluss	Mikrozensuspanel		Beschäftigtenstichprobe	
	n	in %	n	in %
Auszubildende, April 1996	1.504	100,0	24.264	100,0
(n bei gebundener Hochrechnung, in 1.000)	(1.042)		(1.213)	
darunter Erstausbildungen	1.438	95,6	23.066	95,1
- Abgänge aus Status „Auszubildender“ insg. <sup>1)</sup>	459	30,5	8.761	36,1
Erstausbildung mit Lehrabschluss	295	19,6	5.089	21,0
Erstausbildung ohne Lehrabschluss	132	8,8	3.158	13,0
Zweitausbildung	32	2,1	514	2,1
darunter: 1997 ohne Meldung <sup>2)</sup>	102	6,8	1.924	7,9
Erstausbildung mit Lehrabschluss	34	2,3	416	1,7
Erstausbildung ohne Lehrabschluss	59	3,9	1.398	5,8
Zweitausbildung	9	0,6	110	0,5
- Ausfälle / Wegzüge insgesamt	286	19,0		
Erstausbildung	271	18,0		
Zweitausbildung	15	1,0		
+ Zugänge	770	51,2	9.254	38,1
1996 nicht Auszubildender	582	38,7	472	1,9
Zuzug bzw. keine Meldung	188	12,5	8.782	36,2
± Saldo Abgrenzungsänderungen <sup>3)</sup>	+28	1,9	-192	-0,8
= Auszubildende, April 1997	1.557		24.565	
(n bei gebundener Hochrechnung, in 1.000)	(1.077)		(1.228)	

<sup>1)</sup> „Abgänge“ umfassen Wechsel vom Status „Auszubildender“ (April 1996) zum Status „nicht Auszubildender“ (April 1997). Im Mikrozensuspanel ohne Ausfälle. In der Beschäftigtenstatistik einschließlich der Meldelücken zur Berichtswoche im April 1997 (fehlende sozialversicherungspflichtige Beschäftigung bzw. Personen ohne Meldung in der Beschäftigtenstichprobe), jedoch ohne Personen mit letztem Status als Auszubildender.

<sup>2)</sup> Im Mikrozensuspanel beziehen sich die Angaben „ohne Meldung“ auf die Abgänge (ohne Ausfälle), in der Beschäftigtenstichprobe auf die der Meldelücke zur Berichtswoche 4/1997 zeitlich nächsten Meldung.

<sup>3)</sup> Abgrenzungsänderungen beziehen sich auf das Alter und die anerkannten Ausbildungsberufe sowie im Mikrozensuspanel die Zugehörigkeit zur Bevölkerung am Hauptwohnsitz.

Quellen: Mikrozensuspanel 1996-1999 (Bevölkerung am Hauptwohnsitz); Beschäftigtenstichprobe (IABS-R01; Meldungen zu den Mikrozensus-Berichtswochen im April 1996 und 1997); eigene Berechnungen.

Ohne Berücksichtigung der Ausfälle verlassen bis April 1997 von den Auszubildenden im dualen System im April 1996 im Mikrozensuspanel rund 31 Prozent den Ausbildungsstatus, in der Beschäftigtenstichprobe sind es rund 36 Prozent. Die Differenz ist überwiegend auf den in der Beschäftigtenstichprobe höheren Anteil von Abgängen ohne Lehrabschluss zurückzuführen (13 % vs. 9 % im Mikrozensuspanel). Dagegen werden in beiden Datenbasen bei Erstausbildungen jeweils rund 20 Prozent erfolgreicher Lehrabschlüsse ermittelt (19,6 % vs. 21,0 %).

Die Anteile der Stichprobenzugänge unterscheiden sich deutlich zwischen beiden Datenquellen. Im Mikrozensuspanel sind auch Personen enthalten, die erst nach April 1996 eine Ausbildung beginnen. In der Beschäftigtenstichprobe nimmt dagegen die überwiegende Mehrheit der Auszubildenden im April 1997 erstmals eine sozialversicherungspflichtige Beschäftigung auf.

Im Mikrozensuspanel fallen 19 Prozent der Auszubildenden bis 1997 infolge von Wegzügen aus. Sie können weiterhin eine Ausbildung absolvieren, diese abgeschlossen, aber auch abgebrochen haben. Aufgrund des im Mikrozensuspanel im Vergleich zur Beschäftigtenstichprobe geringeren Anteils von Abbrechern ist einerseits zu vermuten, dass ein Teil der im April 1997 ausgefallenen Personen zu den Ausbildungsabbrechern gehört. Andererseits deuten die im Mikrozensuspanel und in der Beschäftigtenstichprobe trotz Ausfällen gleich hohen Abschlussanteile darauf hin, dass erfolgreiche Lehrabsolventen eventuell weniger regional mobil sind als Abbrecher. Diese Vermutungen sollen im Folgenden überprüft werden.

### **6.2.1 Hypothesen zum Ausfall und deskriptive Analysen zum Übergang 1996/97**

Im Anschluss an die bereits im vierten Kapitel dargestellten theoretischen Ansätze zur räumlichen Mobilität bzw. zum Auszug aus dem Elternhaus werden im Folgenden Hypothesen zum Ausfall für die Teilpopulation der Auszubildenden ergänzt und konkretisiert.

In ökonomischen Theorien zur Migration wird die Wanderung als Ergebnis einer Kosten-Nutzen-Rechnung der Akteure gesehen, wobei als Migrationsziel der Ort gewählt wird, der den beruflichen Fähigkeiten am besten entspricht, an dem das höchste Einkommen erzielt werden kann, und für den die direkten und indirekten Kosten (z. B. Umzugskosten, Aufbau neuer Kontakte) am geringsten sind (für einen Überblick siehe Kalter 2000).<sup>61</sup> Jugendliche aus Wohnorten mit einer schwachen wirtschaftlichen Infrastruktur sind häufiger gezwungen, aus Gründen der schulischen oder beruflichen Ausbildung sowie zur Aufnahme einer Erwerbstätigkeit umzuziehen als Personen aus wirtschaftlich prosperierenden Gebieten. Für das östliche Bundesgebiet ist daher eine stärkere Umzugsneigung zu erwarten.

Zwischen dem Übernahmeangebot des Ausbildungsbetriebes bei Abschluss der Ausbildung und Betriebsgröße sowie Wirtschaftszweigen bzw. Berufsgruppen bestehen enge Zusammenhänge (Stegmann und Kraft 1983). Infolge der konjunkturellen Arbeitsmarktprobleme ab Anfang der 1990er Jahre haben sich die Übernahmequoten in allen Betriebsgrößenklassen und Branchen verschlechtert (BMBF 2001: Kap. 4.5; Schöngen 1995: 155; Dietrich und Bellmann 1999). Als Reaktion darauf wurden seit Mitte der 1990er Jahre in einigen Wirtschaftsbereichen tarifliche Regelungen zur Übernahme der Auszubildenden getroffen, die zumindest in der westdeutschen Chemie- und Metallindustrie erfolgreich waren (Bispinck et al. 2002). Aufgrund der zwischen Mikrozensuspanel und Beschäftigtenstichprobe nicht ausreichend vergleichbaren Klassifikationen der Wirtschaftszweige kann diese Fragestellung nicht direkt untersucht werden. Es erscheint aber sinnvoll, in diesem Zusammenhang nach Berufsgruppen zu unterscheiden. Nach Ergebnissen der IAB-Historikdatei für den Zeitraum von 1992 bis 1997 (Haas 2002: 11) weisen Ausbildungsabsolventen in Verkehrs-, Medien-, Körperpflege-, Hotel- und Gaststätten- sowie künstlerischen Berufen eine überdurchschnittliche regionale Mobilität auf.<sup>62</sup> Dabei sind ein Jahr nach Ende der Berufsausbildung in Westdeutschland noch 72 Prozent und in den neuen Bundesländern noch 65 Prozent in ihrem ersten Beschäftigungsbetrieb tätig.

61 Allerdings zeigen neuere Ergebnisse, dass beschäftigungsrelevante Faktoren wie Arbeitslosigkeit weniger stark mit räumlicher Mobilität verbunden sind als aufgrund der ökonomischen Theorie zu erwarten ist (Schachter 2001; Windzio 2004).

62 Die regionale Mobilität bezieht sich auf Arbeitsortwechsel auf Kreisebene.

Ein erfolgreicher Übergang in den Beruf ist mit einem höheren eigenen Erwerbseinkommen verbunden, welches die Auszugsneigung fördert. Andererseits kann aber auch für Auszubildende ohne adäquate Weiterbeschäftigungsmöglichkeiten ein Umzug notwendig werden. Wegen dieser Ambivalenz sind keine eindeutigen Aussagen zur räumlichen Migrationsneigung möglich. Man kann jedoch annehmen, dass erfolgreiche Ausbildungsabsolventen eher ein Übernahmeangebot des Betriebes erhalten als Ausbildungsabbrecher. Insofern bei Betriebswechseln die Wahrscheinlichkeit von Fort- und Umzügen steigt, ist bei Ausbildungsabbrechern mit einer höheren räumlichen Mobilität als bei erfolgreichen Absolventen zu rechnen. Es ist auch davon auszugehen, dass Auszubildende nicht sofort nach dem Übergang ins Erwerbsleben von zu Hause ausziehen, sondern erst nachdem eine sichere Beschäftigung und stabile Einkommenssituation erreicht sind (Huinink und Konietzka 2000: 8).

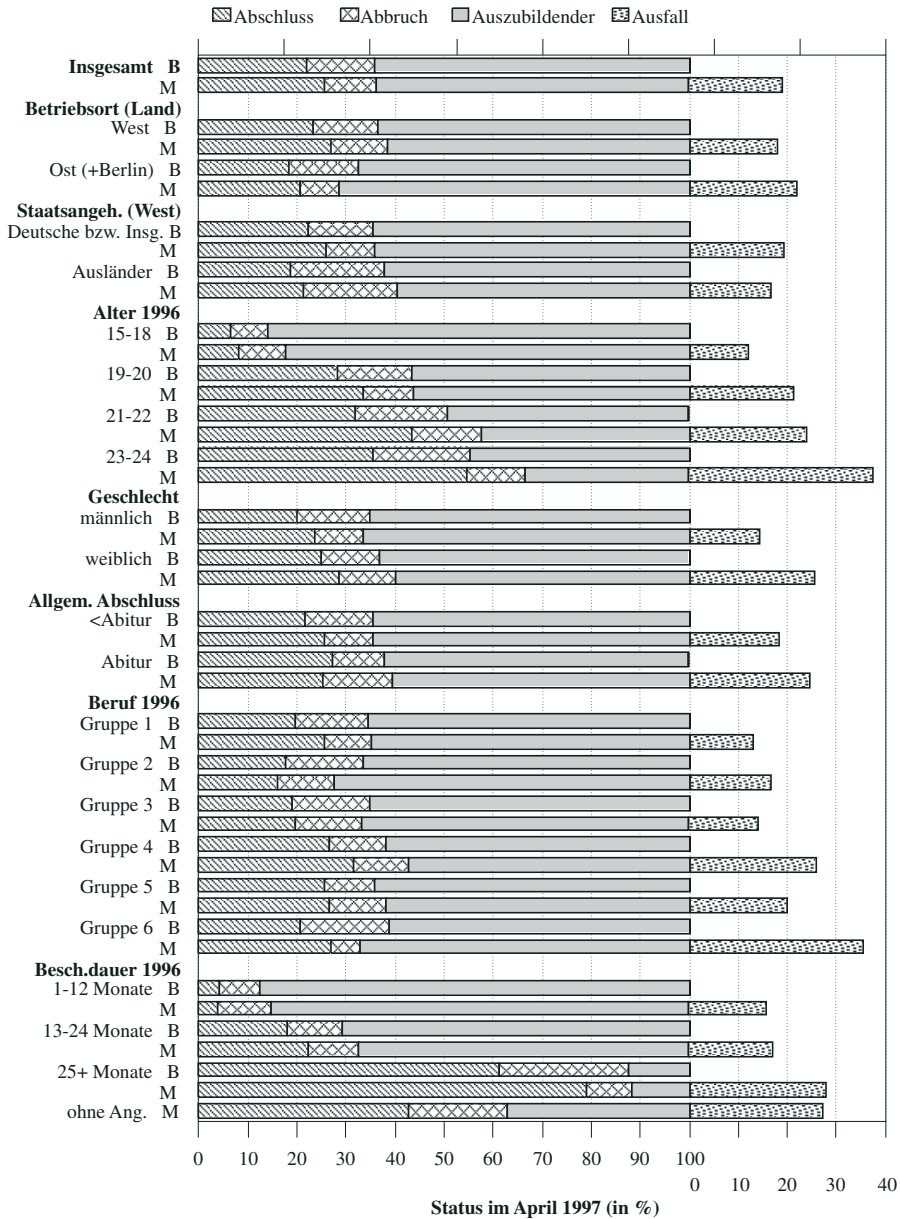
In Bezug auf ethnische Ungleichheiten werden Nachteile von Migranten auf die im Vergleich zu Deutschen geringeren Ressourcen zurückgeführt, die auf dem Arbeitsmarkt nachgefragt werden (Granato und Kalter 2001). Ausländische Jugendliche haben häufiger als deutsche Jugendliche höchstens einen Hauptschulabschluss oder ihren Schulabschluss im Herkunftsland erworben. Ihre Chancen auf einen Ausbildungsplatz sind insbesondere bei geringen Deutschkenntnissen schlechter als bei deutschen Schülern. Eine Studie zu Jugendlichen ohne Berufsausbildung (Troltsch et al. 1999: 70) weist bei Ausländern auf eine im Vergleich zu Deutschen höhere Abbruchquote der beruflichen Ausbildung hin. Stellt man jedoch die Zusammenhänge zwischen dem erreichten allgemeinen Schulabschluss und den Chancen auf einen Ausbildungsplatz in Rechnung, stellen ausländische Auszubildende eine selektive Auswahl dar. Sie dürften sich in Bezug auf den erfolgreichen Abschluss nur wenig von deutschen Jugendlichen unterscheiden (vgl. Seifert und Bender 1996).

Migranten beginnen aufgrund stärker traditionaler Familienkulturen (Nauck 2001) früher als Deutsche mit der Gründung einer eigenen Familie. Geht man aber davon aus, dass eine Familiengründung sowohl bei Migranten als auch bei Deutschen in enger Verbindung mit dem Abschluss der beruflichen Ausbildung steht, ist nicht zu erwarten, dass Ausländer früher als Deutsche aus dem Elternhaus ausziehen.

In welcher Weise die Ausfälle der Auszubildenden mit soziodemografischen Merkmalen und ausbildungsbezogenen Variablen zusammenhängen, zeigt Abbildung 6.2 für das Mikrozensuspanel und die Beschäftigtenstichprobe. Im Unterschied zu Tabelle 6.2 konzentriert sich die Analyse auf Auszubildende in Erstausbildungen.<sup>63</sup> Bei fehlenden Angaben zum Bundesland der Betriebsstätte im Mikrozensuspanel wird das Land des Wohnsitzes verwendet. Die 1996 ausgeübten Ausbildungsberufe sind teilweise schwach besetzt und werden deshalb zu Gruppen zusammengefasst. Während zum Status im April 1997 in der Beschäftigtenstichprobe vollständige Angaben vorliegen, ist dies im Mikrozensuspanel nur für die räumlich Immobilen der Fall. Die Anteile zum Status beziehen sich deshalb im Mikrozensuspanel nur auf räumlich Immobile, die Anteile für den Ausfall beziehen sich auf alle Personen.

63 Personen, die 1996 Auszubildende ohne beruflichen Ausbildungsabschluss waren, aber im April 1997 einen höheren beruflichen Ausbildungsabschluss als die Lehre angegeben haben (z. B. Hochschule), werden als potenzielle Fehlklassifikationen ausgeschlossen.





Quelle: Mikrozensuspanel 1996-1999 (n = 1.429), Beschäftigtenstichprobe (n = 23.044); Auszubildende mit Erstausbildung in anerkannten Ausbildungsberufen 1996; nicht hochgerechnete bzw. ungewichtete Daten; eigene Berechnungen (siehe Tabelle 6.2 und A.4).<sup>64</sup>

Abbildung 6.2: Status der 1996 15- bis 24-jährigen Auszubildenden im April 1997 in der Beschäftigtenstichprobe (B) und im Mikrozensuspanel (M) - in Prozent

64 Beim Mikrozensuspanel beziehen sich die Anteile für Ausfall auf alle Auszubildenden, die anderen Anteile beziehen sich nur auf Personen ohne Ausfall bzw. Wegzug.

Gruppen der Ausbildungsberufe 1996: 1) Pflanzenbauer bis Ernährungsberufe, 2) Bauberufe, 3) Raum-

Abbildung 6.2 zeigt, dass für insgesamt rund 19 Prozent der Auszubildenden im Mikrozensuspanel infolge von Wegzügen 1997 keine Informationen vorliegen (siehe auch Tab. A4 im Anhang). Der Vergleich räumlich immobiler Personen des Mikrozensus zur Gesamtheit der Beschäftigtenstichprobe zeigt Folgendes: Die Anteile der Personen, die sich 1997 noch in der dualen Ausbildung befanden, sind mit jeweils rund 64 Prozent gleich. Dagegen sind Lehrabsolventen im Mikrozensuspanel (25,5 %) im Vergleich zur Beschäftigtenstichprobe (22,1 %) etwas über- und Ausbildungsabbrecher (10,6 % vs. 13,6 %) unterrepräsentiert.

Aufgrund des in der Beschäftigtenstichprobe begrenzten Variablenumfangs können von den oben genannten Faktoren des Auszugsverhaltens nicht alle Aspekte untersucht werden. Die nur mit dem Mikrozensuspanel durchführbaren aber hier nicht weiter dokumentierten Ausfallanalysen zu zwei Annahmen im Bereich der subjektiven Handlungsstrukturen (siehe Abschnitt 4.1) zeigten, dass Auszubildende, die 1996 nicht bei beiden Eltern wohnten, in höherem Maße räumlich mobil sind bzw. bis 1997 ausfallen (24 %) als Auszubildende, die bei beiden Eltern lebten (13 %). Des Weiteren fallen Auszubildende in Kernstädten überdurchschnittlich häufig aus (28 % vs. 19 % insgesamt).

Wie zu erwarten, finden sich überdurchschnittliche Ausfälle, wenn die Ausbildungsstätte 1996 in Ostdeutschland (einschließlich Berlin) lag. Für beide Gebiete zeigt sich, wie für das Bundesgebiet insgesamt, dass räumlich immobile Absolventen im Mikrozensuspanel im Vergleich zur Beschäftigtenstichprobe übererfasst und räumlich immobile Abbrecher untererfasst sind. Die Untererfassung der Abbrecher ist jedoch in Ostdeutschland mit rund sechs Prozentpunkten Differenz zur Beschäftigtenstichprobe deutlich stärker ausgeprägt als in Westdeutschland, wo sie nur zwei Prozent beträgt. Diese Unterschiede deuten darauf hin, dass in Ostdeutschland Ausbildungsabbrecher, vermutlich aufgrund der dort prekären Arbeitsmarktsituation, räumlich mobiler sind als erfolgreiche Absolventen in Ost- und Westdeutschland.

Bei der weiteren Überprüfung dieser Frage, in welcher Weise Statuswechsel und Ausfälle zusammenhängen, ist man mit dem Problem konfrontiert, dass in der Beschäftigtenstichprobe lediglich Wechsel des Betriebsortes feststellbar sind, ein anderer Arbeitsort aber auch ohne Änderung des Wohnsitzes, z. B. durch Pendeln, erreicht werden kann. Eine Auswertung der Beschäftigtenstichprobe ergibt, dass von allen 1996 in westdeutschen Betrieben beschäftigten Auszubildenden 96 Prozent der Absolventen und 93 Prozent der Abbrecher 1997 im gleichen Bundesland arbeiten. Dagegen sind von allen 1996 in ostdeutschen Betrieben beschäftigten Auszubildenden ein Jahr später 90 Prozent der Absolventen und 87 Prozent der Ausbildungsabbrecher hinsichtlich des Bundeslandes der Betriebsstätte immobil. Von den 1996 in ostdeutschen Betrieben beschäftigten Absolventen und Abbrechern sind jeweils etwa die Hälfte der räumlich Mobilen zu Betrieben im westdeutschen Bundesgebiet gewechselt.<sup>65</sup> Die auf das Bundesland des Betriebsortes bezogenen Unterschiede der räumlichen Mobilität zwischen West und Ost sowie zwischen Absolventen und Abbrechern sind gering und damit nur ein schwaches Indiz für die im Mikrozensuspanel beobachteten unterschiedlichen Ausfallmuster.

---

ausstatter bis Techniker, 4) Waren- und Dienstleistungskaufleute, 5) Organisations- bis Erziehungsberufe, 6) Allgemeine Dienstleistungsberufe und Verkehrsberufe.

Im Mikrozensuspanel liegen zur Beschäftigungsdauer beim gegenwärtigen (1996) Arbeitgeber für drei Prozent ( $n = 48$ ) keine Angaben vor. Die Angaben zur Beschäftigtenstichprobe beziehen sich auf die Betriebsstätte im April 1996.

65 Im Zeitverlauf hat die räumliche Mobilität in Ostdeutschland deutlich zugenommen. Im Übergang 1998/99 wechseln 13 Prozent der Ausbildungsabsolventen, die 1998 in ostdeutschen Betriebsstätten arbeiteten, das Bundesland des Betriebsortes, neun Prozent davon zu Betrieben in Westdeutschland.



In Bezug auf die 1996 ausgeübten Ausbildungsberufe weisen im Mikrozensuspanel Auszubildende in Agrar- bis Ernährungsberufen (Gruppe 1) unterdurchschnittliche Ausfälle von 13 Prozent auf. Überdurchschnittliche Ausfälle sind bei den Berufsgruppen der Waren- und Dienstleistungskaufleute (Gruppe 4: 26 %) sowie der Verkehrs- bzw. allgemeinen Dienstleistungsberufe (Gruppe 6: 36 %) festzustellen. Die Berufsgruppe 4 konzentriert sich zu einem Drittel auf wirtschaftsbezogene Dienstleistungen, die wie das Kredit- und Versicherungsgewerbe überdurchschnittliche Übernahmequoten von Auszubildenden in ein Beschäftigungsverhältnis aufweisen (BMBF 1999: 147). Auszubildende der Berufsgruppe 6 sind nach Auswertungen der Beschäftigtenstatistik überwiegend im Bereich haushaltsbezogener Dienstleistungen beschäftigt, der eine unterdurchschnittliche Übernahmequote besitzt. Stärker als bei den anderen Berufsgruppen sind in der Berufsgruppe 6 Ausbildungsabbrecher des Mikrozensuspanels gegenüber der Beschäftigtenstatistik mit zwölf Prozent untererfasst. Zwar ist diese Gruppe schwach besetzt ( $n = 104$ ), dennoch kann die Untererfassung der Abbrecher als Indiz dafür verstanden werden, dass Ausfälle bzw. Wegzüge mit Statuswechseln (Ausbildungsabschluss und -abbruch) beim Berufseinstieg zusammenhängen.

Abbildung 6.2 zeigt für das Geschlecht, den allgemeinen Schulabschluss und die Staatsangehörigkeit die erwarteten Zusammenhänge. Frauen sind mit 26 Prozent deutlich räumlich mobiler als Männer (14 %). Überdurchschnittlich hoch ist auch die Ausfallquote von Auszubildenden mit Fachhochschulreife oder Abitur (25 %). Hinsichtlich der Staatsangehörigkeit, die nur für das westliche Bundesgebiet ausgewiesen wird, sind die Ausfallquoten von Ausländern und Deutschen im Mikrozensuspanel nahezu gleich.<sup>66</sup> Ausländische Auszubildende brechen aber im Vergleich zu Deutschen häufiger die Ausbildung ab.

Von allen in Abbildung 6.2 betrachteten Merkmalen hängt der Ausfall bzw. Wegzug im Mikrozensuspanel am stärksten mit dem Alter und der bis 1996 erreichten Beschäftigungsdauer im Ausbildungsbetrieb zusammen. Während von den im April 1996 unter 19 Jahre alten Auszubildenden ein Jahr später zwölf Prozent weggezogen sind, sind es bei den 21- bis 22-Jährigen 24 Prozent, bei den 23- bis 24-Jährigen sogar 37 Prozent.

Bei einer Beschäftigungsdauer<sup>67</sup> von mehr als zwei Jahren im Ausbildungsbetrieb liegt die Ausfallquote mit 28 Prozent um zehn Prozentpunkte höher als bei einer kürzeren Beschäftigungsdauer. Sowohl in Bezug auf das Alter als auch hinsichtlich der Beschäftigungsdauer weichen die Anteile der Absolventen und Abbrecher des Mikrozensuspanels von den Anteilen der Beschäftigtenstichprobe ab. So liegen die Abschlussanteile für die beiden oberen Alterskohorten der 21- bis 24-Jährigen um zwölf bzw. 19 Prozentpunkte im Mikrozensuspanel höher als in der Beschäftigtenstichprobe. Bei einer Beschäftigungsdauer von über zwei Jahren sind räumlich immobile Ausbildungsabsolventen im Mikrozensuspanel im Vergleich zur Gesamtheit in der Beschäftigtenstichprobe um 18 Prozentpunkte übererfasst (79 % vs. 61 %), Ausbildungsabbrecher sind dagegen um 17 Prozentpunkte untererfasst (9 % vs. 26 %). Das Muster der Verteilungsabweichungen zur Beschäftigtenstichprobe korreliert mit der Ausfallrate des Mikrozensuspanels, d. h. je höher die Ausfallrate, desto stärker sind im Mikrozensuspanel räumlich immobile Absol-

66 Eine weitere Unterscheidung von Ausländern nach Migrationsmerkmalen wie Herkunftsland oder Einreisalter ist in der Beschäftigtenstichprobe nicht möglich.

67 Die Ausbildungsdauer ist im Mikrozensus nicht direkt verfügbar. Es wird aber nach dem Beginn der gegenwärtigen Tätigkeit beim derzeitigen Arbeitgeber gefragt. Diese Angaben stimmen zwar (z. B. beim Wechsel des Ausbildungsbetriebes) nicht mit dem „Lehrjahr“ überein, sie eignen sich jedoch gut als Proxyvariable. Für die Beschäftigtenstichprobe wird die Dauervariable mithilfe des Vergleichs der Nummern der Betriebsstätten im Beschäftigungsverlauf konstruiert. Die Randverteilungen der Beschäftigungsdauern beider Datensätze unterscheiden sich nur geringfügig um maximal drei Prozent.

venten über- und Ausbildungsabbrecher untererfasst. Dies deutet, wie bereits bei den Merkmalen Betriebsort und Beruf angesprochen, darauf hin, dass erfolgreiche Ausbildungsabsolventen räumlich immobiler sind als Ausbildungsabbrecher. Damit wären Ausfälle im Mikrozensuspanel direkt mit dem interessierenden Statuswechsel verbunden. Darüber hinaus ist zu erwarten, dass diese Zusammenhänge unter anderem von regionalen Kontexten abhängen und mit weiteren Merkmalen korreliert sind. Für die Untersuchung von Kovariatenzusammenhängen mit der Auszugsrate und um den Einfluss von Drittvariablen zu kontrollieren, werden im Folgenden log-lineare Modelle herangezogen.

Wie bei den meisten statistischen Modellen werden dabei von Messfehlern freie Variablen angenommen (vgl. allgemein dazu Skrondal und Rabe-Hesketh 2004; vgl. zu Verzerrungen aufgrund von Messfehlern bei Übergangsanalysen Bassi et al. 2000). Jedoch wurde im Zusammenhang mit den Verteilungsabweichungen zwischen Mikrozensuspanel und Beschäftigtenstichprobe vermutet, dass im Mikrozensuspanel Ausbildungsabsolventen eventuell als Folge der Fokussierung auf den Interviewzeitpunkt bei der Frage nach dem Schulbesuch übererfasst sind (siehe Seite 118). Falls dies zutrifft, kann in den statistischen Modellen nicht mehr eindeutig zwischen Ausfalleffekten und Klassifikationsfehlern unterschieden werden. Die Überprüfung dieser Vermutung ist allerdings schwierig, weil für den Mikrozensus keine Information über den Interviewzeitpunkt vorhanden ist. Behelfsweise wird die im April 1996 erfragte Dauer der Beschäftigung beim damaligen Arbeitgeber als Kontrollvariable verwendet. Bei Fokussierungseffekten im Mikrozensuspanel ist bei Ausbildungsabsolventen eine kürzere Beschäftigungsdauer als in der Beschäftigtenstichprobe zu erwarten.

Die in Tabelle 6.3 berichteten Kennwerte zeigen jedoch weitestgehend gleiche Verteilungen der Beschäftigungsdauern von räumlich Immobilen des Mikrozensuspanels und der Beschäftigtenstichprobe. Für die zum Zeitpunkt April 1997 ermittelten Absolventen beträgt die mittlere Beschäftigungsdauer (Median) 32 Monate. Die Mediane der Folgejahre liegen jeweils rund zwölf Monate niedriger. Zwar können mit dieser einfachen Überprüfung Fokussierungseffekte nicht gänzlich ausgeschlossen werden, doch spricht das Ergebnis dafür, dass die vermuteten Klassifikationsfehler in den folgenden Ausfallanalysen nicht auftreten.

**Tabelle 6.3:** Perzentile und arithmetische Mittelwerte der Beschäftigungsdauer im Betrieb (Stand: April 1996; in Monaten) für Auszubildende des dualen Systems im April 1996 mit Ausbildungsabschluss bis 1999 im Mikrozensuspanel und in der Beschäftigtenstichprobe

Ermittelter Abschluss	Kennwert	Mikro- zensuspanel	Beschäftigten- stichprobe
4/1997	Absolventen (n)	279	4.673
	Anteil (% v. Auszubildenden 4/96)	20,6	20,3
	Beschäftigungsdauer: 1. Quartil	21,0	21,2
	Median	32,0	32,3
	3. Quartil	33,0	33,4
	Mittelwert	26,7	28,0
4/1998	Absolventen (n)	241	4.782
	Anteil (%)	17,8	20,7
	Beschäftigungsdauer: 1. Quartil	9,0	9,0
	Median	20,0	20,2
	3. Quartil	21,0	21,2
	Mittelwert	17,2	17,4
4/1999	Absolventen (n)	174	3.523
	Anteil (%)	12,9	15,3
	Beschäftigungsdauer: 1. Quartil	8,0	8,0
	Median	9,0	9,0
	3. Quartil	9,0	9,0
	Mittelwert	9,6	9,2

*Quellen:* Mikrozensuspanel 1996-1999 (nur räumlich immobile Absolventen); IAB-Beschäftigtenstichprobe (ohne Abschlüsse bei Meldelücken); ungewichtete Daten; eigene Berechnungen (siehe Anmerkungen zu Abbildung 6.2).

## 6.2.2 Modellergebnisse zum Ausfall im Mikrozensuspanel

Wie in Abschnitt 3.3.3 ausgeführt, bietet das log-lineare Modell zwar Möglichkeiten, Überlegungen zu testen, wie Ausfälle und Analysevariablen zusammenhängen, die Gültigkeit der Ergebnisse hängt dabei allerdings von den Modellannahmen ab. Für die im Mikrozensuspanel ausgefallenen Auszubildenden des Jahres 1996 liegen keine Informationen zu ihrem bis April 1997 erreichten Status vor. Deshalb kann ohne zusätzliche Informationen zum Beispiel die MAR-Annahme nicht kontrolliert werden, also ob sich die Gruppen der vor Ermittlung des Status weggezogenen und nicht weggezogenen Auszubildenden hinsichtlich der bei der Ausfallmodellierung berücksichtigten Variablen systematisch unterscheiden. Zur Validierung von Modellen, die einen nicht ignorierbaren Ausfallprozess annehmen, sind fundierte a priori Kenntnisse notwendig (Allison 2002: 4-5). Im Folgenden wird die Beschäftigtenstichprobe als externe Datenquelle zur Validierung herangezogen. Im Unterschied zu der im Unterabschnitt 3.3.4 beschriebenen Vorgehensweise werden weitere erklärende Variablen verwendet.

Für das Mikrozensuspanel ist anzunehmen, dass Zugezogene und Fortgezogene gleiche Eigenschaften aufweisen. Grundsätzlich bietet deshalb die Berücksichtigung von 1997 zugezogenen Auszubildenden gute Möglichkeiten für die Schätzung von Ausfallmodellen. Jedoch liegen Angaben zur Differenzierung der Zugezogenen nach ihrem Status als Auszubildender im Vorjahr

nur in der Unterstichprobe zur Arbeitskräfteerhebung (0,45 %) vor. Da nur sehr wenige Zugezogene, die 1996 eine duale Ausbildung absolvierten, identifizierbar sind, müssen im Folgenden Angaben von Zugezogenen außer Acht bleiben.<sup>68</sup>

Die Analysen beschränken sich daher auf die bereits in den vorigen Abschnitten beschriebenen Auszubildenden des Jahres 1996 in Erstausbildung im Alter bis zu 24 Jahren. Ausgeschlossen werden 48 Personen, für die zur Beschäftigungsdauer beim Arbeitgeber (1996) im Mikrozensuspanel keine Angaben vorliegen. In der abhängigen Variablen (Status 1997;  $Y$ ) werden drei Kategorien unterschieden: (1) Auszubildende im April 1997, (2) Personen, bei denen 1997 kein Lehrabschluss ermittelt werden konnte („Abbrecher“) sowie (3) Auszubildende mit Angaben zum erfolgreichen Lehrabschluss im April 1997 („Abschluss“). Für gültige Werte in  $Y$  ist der Ausfallindikator  $R$  gleich Eins. Die bis 1997 weggezogenen Auszubildenden werden im Ausfallindikator mit  $R = 0$  codiert. Um Schätzprobleme durch zu kleine Zellenbesetzungen zu vermeiden, werden die im Unterabschnitt 6.2.1 verwendeten Zusammenfassungen übernommen. Tabelle 6.4 enthält die Ergebnisse der log-linearen Modelle zum Status 1997 und zum Ausfall.

Die in Spalte 2 der Tabelle 6.4 dargestellten Modelle geben die Logit-Spezifikation des zugrunde liegenden log-linearen Modells wieder. Der erste Block bis zum Modell 9 zeigt Ergebnisse zum strukturellen Zusammenhang zwischen dem Status und den erklärenden Variablen, die unter der Annahme eines vollständig zufälligen Ausfalls (MCAR) geschätzt werden.<sup>69</sup> Das zweite Modell, in dem Zusammenhänge aller erklärenden Variablen mit dem Status angenommen werden, stellt für die Frage nach dem partiellen Effekt der einzelnen erklärenden Variablen das Basismodell dar.

Es ist zu erwarten, dass die Übergänge von Auszubildenden des Jahres 1996 zur Fortsetzung der Ausbildung in 1997 ( $Y=1$ ), dem erfolgreichen Lehrabschluss ( $Y=2$ ) oder der Beendigung des Ausbildungsverhältnisses ohne Abschluss ( $Y=3$ ) maßgeblich vom Fortschritt der bisherigen Ausbildung abhängen. Dementsprechend stellt sich in den Vergleichen der Modelle 3 bis 9 mit dem Basismodell (#2) bzw. den entsprechenden Pseudo- $R^2$ -Werten die Beschäftigungsdauer (Variable  $T$ ) als dominante Einflussgröße (Modell 9 vs. 2) heraus. Dies sowie die weiteren Modellergebnisse zu den strukturellen Zusammenhängen korrespondieren mit den obigen deskriptiven Befunden (vgl. Abb. 6.2), sodass auf eine Diskussion verzichtet werden kann.

68 Bei der Frage nach der Stellung im Beruf des Vorjahres sind Schüler, Studenten und Auszubildende in einer gemeinsamen Antwortkategorie zusammengefasst und somit nicht ausreichend zu unterscheiden. Die Teilnahme an einer beruflichen Aus- und Weiterbildung seit dem Vorjahr wird ohne Auskunftspflicht erfragt. Informationen zu dem 1996 eingenommenen Status erhält man nur für jene Personen, die 1997 („gegenwärtig“) nicht an einer beruflichen Aus- und Fortbildung teilgenommen haben. Auswertungen ergaben, dass von den 1997 Zugezogenen ( $n = 3.706$ ) für nur knapp fünf Prozent ( $n = 176$ ) die Teilnahme an einer Aus- und Weiterbildung im Vorjahr angegeben ist. Von diesen Teilnehmern können lediglich neun Prozent als Auszubildende des dualen Systems eingestuft werden.

69 In den Modellen wird von einer uneingeschränkten Zufallsstichprobe ausgegangen. Das Mikrozensuspanel ist dagegen eine mehrfach geschichtete Klumpenstichprobe. Die Vernachlässigung der Klumpung ist i. d. R. mit einer Unterschätzung des Standardfehlers der Koeffizienten verbunden. Verwendet man zur Berücksichtigung der Klumpung eine robuste Varianzschätzung (siehe Rogers 1993; Wooldridge 2002: 401ff.), unterscheiden sich die Ergebnisse von Logit-Modellen jedoch nicht von Modellen, die unter der Annahme unabhängig identisch verteilter Beobachtungen geschätzt wurden.

Tabelle 6.4: Log-lineare Modelle zum Status ( $Y$ ) und Ausfall ( $R$ ) 1997 der bis 24-jährigen Auszubildenden des Jahres 1996 (Erstausbildungen)

#	Logit-Spezifikation	Log-	Likelihood-Ratio-Statistik				R² (in %)	
		Likelihood	df	G²	df	P	AIC	YIX RI...
MCAR-Modelle								
1	(Y) (RD)	-151.658,11	3				777	
2	#1 + Y * (D, A, G, L, B, S, J, T)	-145.865,32	31	11.585,58	28	0,00	-10.752	27,0
3	#2 – YA [A]lterskohorte]	-146.129,29	25	527,94	6	0,00	-10.236	25,8
4	#2 – YG [G]eschlecht]	-145.869,43	29	8,20	2	0,02	-10.748	27,0
5	#2 – YL [L]and]	-145.910,64	29	90,63	2	0,00	-10.666	26,8
6	#2 – YB [B]ildung]	-145.889,93	29	49,21	2	0,00	-10.707	26,9
7	#2 – YS [S]taatsangeh.]	-145.897,86	29	65,08	2	0,00	-10.691	26,8
8	#2 – YJ [J]ob/Berufsggr.]	-146.002,10	21	273,55	10	0,00	-10.499	26,4
9	#2 – YT [T] Beschäftigt.]	-149.857,91	27	7.985,17	4	0,00	-2.775	8,4
MAR-Modelle								
10	#2 + RD * (A, G, L, B, S, J, T)	-145.819,36	45	91,92	14	0,00	-10.816	27,0 3,2
11	#10 – RDA	-145.829,51	42	20,29	3	0,00	-10.802	27,0 2,5
12	#10 – RDG	-145.824,74	44	10,75	1	0,00	-10.807	27,0 2,8
13	#10 – RDL	-145.822,14	44	5,56	1	0,02	-10.813	27,0 3,0
14	#10 – RDB	-145.819,42	44	0,11	1	0,73	-10.818	27,0 3,2
15	#10 – RDS	-145.819,39	44	0,05	1	0,83	-10.818	27,0 3,2
16	#10 – RDJ	-145.828,78	40	18,83	5	0,00	-10.807	27,0 2,6
17	#10 – RDT	-145.822,41	43	6,08	2	0,05	-10.814	27,0 3,0
18	#2 + RD * (A, G, L, J, T)	-145.819,44	43	0,15	2	0,93	-10.820	27,0 3,2
NINR-Modelle								
19	#18 + RDY	-145.802,94	45	32,99	2	0,00	-10.849	26,9 6,8
20	#19 + RDTY <small>[Y*: 1=2 vs. 3]</small>	-145.798,37	47	9,15	2	0,01	-10.854	27,0 7,4

\*) Unter „YIX“ sind die Pseudo-R<sup>2</sup>-Werte für das Strukturmodell, unter „RI...“ die Werte für das Ausfallmodell zu finden. Da in den hier verwendeten Daten der Beschäftigtenstichprobe (Datentyp  $D = 1$ ) keine Ausfälle ( $R = 0$ ) vorkommen, wird im Ausfallmodell grundsätzlich die Interaktion  $RD$  mit der Definition dieser strukturellen Nullzelle ( $D = 1$  &  $R = 0$ ) berücksichtigt. Das R<sup>2</sup> beträgt für die Modelle 1-9 bereits 53,5 Prozent, wird aber in der Spalte „RI...“ auf Null gesetzt und nicht berichtet. Die ab Modell 10 genannten Werte entsprechen der Differenz zwischen dem jeweils ermittelten R<sup>2</sup> zum Basiswert von 53,5 Prozent. (Siehe Tabelle A.5)

Ausgehend von den im Modell 2 berücksichtigten strukturellen Zusammenhängen werden in den MAR-Modellen (10-18) zusätzlich Zusammenhänge zwischen dem Ausfall ( $R$ ) und den erklärenden Variablen geprüft. Nach Berücksichtigung der in der Beschäftigtenstichprobe vorliegenden strukturellen Nullzellen (s. o.) werden die Interaktionen zwischen Ausfall, Datentyp und erklärenden Variablen geschätzt. Wie der AIC-Wert zum Modell 10 zeigt, führt dies insgesamt zu einer deutlichen Modellverbesserung. Die Annahme vollständig zufälliger Ausfälle trifft damit nicht zu.<sup>70</sup> Die Rückwärtsselektionen zur Ermittlung der partiellen Effekte zeigen für die Varia-

70 Bei ausschließlich diskreten Daten können die Koeffizienten eines Strukturmodells und die Koeffizienten des Ausfallmodells nicht vollständig unabhängig geschätzt werden (Vermunt 1997a: 77), sodass

blen Bildungsabschluss (Variable *B*; Modell 14) und Staatsangehörigkeit (Variable *S*; Modell 15) keine signifikanten Zusammenhänge. Alle anderen Variablen interagieren insgesamt statistisch signifikant mit dem Ausfall, wobei die jeweiligen Koeffizienten die bereits in der deskriptiven Analyse erkennbaren Zusammenhänge widerspiegeln (siehe Tabelle A.5).

In Bezug auf die partiellen Erklärungsleistungen hängt der Ausfall im Mikrozensuspanel mit einem partiellen Pseudo- $R^2$  von jeweils rund 0,7 Prozent am stärksten mit dem Alter (Variable *A*; Modell 11 vs. 10: 3,2 % - 2,5 %) und dem Ausbildungsberuf (Variable *J*; Modell 16 vs. 10: 3,2 % - 2,6 %) zusammen. Bei den Ausbildungsberufen unterscheiden sich allerdings nur die kaufmännischen und Dienstleistungsberufe (Gruppe 4 und 6) statistisch signifikant von der Referenzgruppe (Pflanzenbauer bis Ernährungsberufe).

Die im Strukturmodell dominierende Variable Beschäftigungsdauer (*T*; Modell 17) nimmt bei den partiellen Effekten nach dem Geschlecht (Modell 12) erst den vierten Rang ein. Die Ausfallneigung von Auszubildenden mit einer Beschäftigungsdauer von 13 bis 24 Monaten ist nicht signifikant von der von Personen mit einer kürzeren Dauer verschieden (siehe Tabelle A.5). Entfernt man die Interaktionen der nicht signifikanten Variablen, lässt sich das Ausfallgeschehen mit dem Modell 18 sparsamer abbilden.<sup>71</sup>

In den deskriptiven Analysen zeigten sich in Form der Übererfassung von Absolventen und der Untererfassung von Ausbildungsabbrechern des Mikrozensus im Vergleich zur Gesamtheit der Beschäftigtenstichprobe Hinweise darauf, dass der Abbruch der Ausbildung mit einem höheren Ausfallrisiko einhergeht als der erfolgreiche Abschluss. Ob der Ausfall nicht nur mit den im MAR-Modell 18 berücksichtigten erklärenden Variablen, sondern darüber hinaus mit dem nicht beobachteten Status zusammenhängt, wird mit dem Modell 19 getestet. Wie anhand der Likelihood-Ratio-Statistik und des AIC-Wertes zu sehen ist, kann durch die zusätzlich geschätzte Interaktion zwischen Status und Ausfall im Mikrozensuspanel (*RDY*) eine signifikante Modellverbesserung gegenüber dem Modell 18 erreicht werden.

Mit dem Modell 19 wird im Teilmodell zum Ausfall für Absolventen im Vergleich zur Referenzgruppe der Auszubildenden eine statistisch signifikant niedrigere Ausfallneigung geschätzt ( $\beta_{022}^{RYD} = -2,15$ ; siehe Tabelle A.5). Dagegen unterscheiden sich Abbrecher ( $\beta_{032}^{RYD} = 0,40$ ) nicht signifikant von Auszubildenden, jedoch von Absolventen. Durch das Modell 19 wird ein wesentlich stärkerer Effekt zum Zusammenhang zwischen dem Ausfall und der Beschäftigungsdauer bei mehr als zwei Jahren geschätzt ( $\beta_{023}^{RDT} = 1,51$ ) als durch das MAR-Modell 18 ( $\beta_{023}^{RDT} = 0,44$ ). Die weitere Aufklärung dieser Zusammenhänge wird dadurch erschwert, dass bei einem Modell, das neben der Interaktion zwischen Status und Ausfall (*RDY*) des Modells 19 auch eine Interaktion mit der Beschäftigungsdauer (*RDYT*) enthält, die Koeffizienten zum Abschluss ( $Y=2$ ) nicht identifiziert werden können bzw. eine „boundary solution“ vorliegt. Für eine Beschäftigungsdauer von 13 bis 24 Monaten werden für die Ausfallgruppe durch dieses Modell keine Absolventen geschätzt. Setzt man diese Gruppe in Bezug auf den Ausfall mit der Referenzgruppe der Auszubildenden gleich (*RDTY'*), lässt sich mit dem entsprechenden Modell 20 gegenüber dem Modell 19 eine weitere Verbesserung der statistischen Erklärungsleistung erzielen. Die Koeffizienten sprechen für eine deutlich größere Ausfallneigung von Ausbildungsabbrechern bei einer Beschäftigungsdauer von mehr als zwei Jahren im Vergleich zu Auszubildenden und Absolventen.

---

streng genommen die MAR-Modelle bereits zu den NINR-Modellen zählen. Da aber die Koeffizienten zu den inhaltlich interessierenden Analysevariablen bei den MCAR- und MAR-Modellen stabil sind (siehe Tabelle A.4), werden die idealtypischen Bezeichnungen dennoch verwendet.

71 Weitere Tests ergaben, dass eine Aufnahme höherer Interaktionen der unabhängigen Variablen im Ausfallmodell, z. B. nach dem Land der Betriebsstätte, keine signifikante Modellverbesserung bietet. D. h. es finden sich keine Hinweise auf bedeutsame Ost-/Westunterschiede.



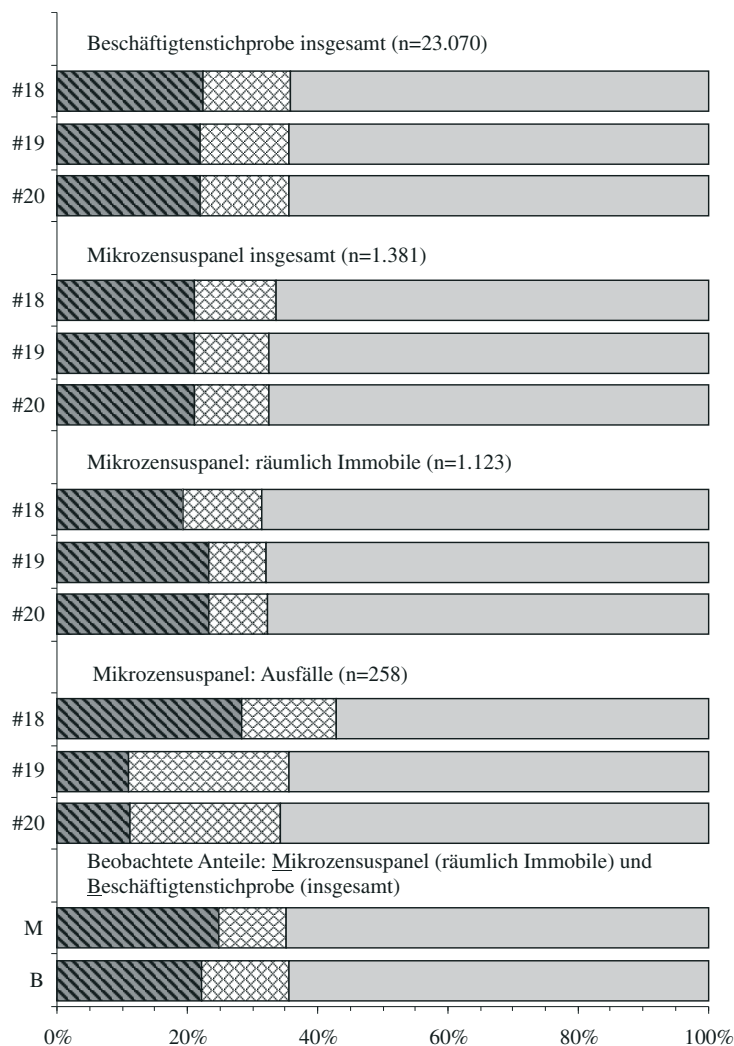
Allerdings erlauben die Koeffizienten des log-linearen Modells keine direkte Einsicht zum Zusammenhang zwischen den inhaltlich interessierenden Variablen für die Gesamtheit einschließlich der Ausfälle. Mithilfe der unter den jeweiligen Modellannahmen geschätzten Anteile der abhängigen Variablen wird es aber möglich einzuschätzen, wie plausibel die in den Modellen 18 bis 20 völlig verschiedenen Annahmen sind.

Betrachtet man in Abbildung 6.3 die Schätzung für die gesamten Stichproben, d. h. beim Mikrozensuspanel einschließlich der Ausfälle, sind praktisch keine modellspezifischen Unterschiede festzustellen. Für das Mikrozensuspanel werden durch das Modell 18 12,5 Prozent Abbrecher, durch die Modelle 19 und 20 mit 11,6 Prozent etwas weniger Abbrecher geschätzt. Entsprechende Verschiebungen finden sich bei den Anteilen der Auszubildenden. Die Unterschiede zwischen Beschäftigtenstichprobe und Mikrozensuspanel insgesamt liegen bei maximal drei Prozent. Berücksichtigt man den Stichprobenfehler, können Abweichungen in dieser Größe als vernachlässigbar betrachtet werden.

Bei der Interpretation von unterschiedlichen Schätzergebnissen sind die Modellannahmen zu beachten. Für das Mikrozensuspanel und die Beschäftigtenstichprobe werden die gleichen strukturellen Zusammenhänge zwischen dem Status ( $Y$ ) und den erklärenden Variablen angenommen. Wegen der im Vergleich zum Mikrozensuspanel erheblich höheren Fallzahl der Beschäftigtenstichprobe spiegeln die Schätzungen vor allem die Zusammenhangsstrukturen der Beschäftigtenstichprobe wider. Dies ist der Grund, weshalb für die räumlich Immobilen im Mikrozensuspanel die geschätzten Anteile (Zeilen 7-9 in Abb. 6.3) von den für diese Gruppe beobachteten Anteilen (Zeile 13) abweichen. Da für das Mikrozensuspanel die geschätzte Verteilung der abhängigen Variablen nicht durch eine Interaktion mit dem Datentyp restringiert ist, bietet sich für die Modellevaluation ein Vergleich der geschätzten und beobachteten Werte an. Zu präferieren ist dabei ein Modell, mit dem die beobachteten Randverteilungen möglichst gut repliziert werden.

Für das MAR-Modell 18 betragen die Abweichungen zwischen geschätzten und beobachteten Anteilen beim Abschluss -5,5, beim Abbruch 1,7 und bei Auszubildenden 3,9 Prozentpunkte. Dagegen liegen die Schätzungen der NINR-Modelle 19 und 20 deutlich näher bei den beobachteten Anteilen. Die Differenzen betragen bei den Kategorien Abschluss und Abbruch jeweils nur rund 1,5 und bei Auszubildenden drei Prozentpunkte. Somit spricht die beim MAR-Modell 18 im Vergleich zu den NINR-Modellen für räumlich immobile Personen schlechtere Anpassung der geschätzten Anteile an die beobachteten Anteile auch gegen die Annahme bedingt zufälliger Ausfälle.

Insbesondere für die ausgefallenen Personen im Mikrozensuspanel zeigt Abbildung 6.3 deutlich die Konsequenzen der verschiedenen Annahmen zum Ausfall. Im Vergleich zu räumlich Immobilen werden unter der Annahme bedingt zufälliger Ausfälle durch das MAR-Modell 18 für ausgefallene Personen (Zeile 10 in Abb. 6.3) fast gleiche Anteile von Ausbildungsabbrüchen (Ausfälle: 15 %; Immobile: 12 %), aber wesentlich höhere Anteile von Abschlüssen (Ausfälle: 28 %; Immobile: 19 %) und geringere Anteile von Auszubildenden (Ausfälle: 57 %; Immobile: 69 %) geschätzt. Dagegen werden mit den NINR-Modellen, in denen ein Zusammenhang zwischen dem Ausfall und dem nicht beobachteten Status angenommen wird, die Anteile der Auszubildenden unter den Ausgefallenen ähnlich wie bei den räumlich Immobilen geschätzt. Die Prozentsatzverhältnisse der für den Abschluss und Abbruch geschätzten Anteile des MAR-Modells 18 (Abschluss 28 %/Abbruch 15 % = 2) kehren sich bei den NINR-Modellen 19 (11 %/24 % = 0,4) und 20 (11 %/23 % = 0,5) näherungsweise um.



Modellterme (siehe Tabelle 6.4):

Modell #18  $Y * (D, A, G, L, B, S, J, T) + RD * (A, G, L, J, T)$

#19  $Y * (D, A, G, L, B, S, J, T) + RD * (A, G, L, J, T) + RDY$

#20  $Y * (D, A, G, L, B, S, J, T) + RD * (A, G, L, J, T) + RDY + RDTY$   
 [Restriktion bei RDTY für Y: 1=2 vs. 3]

Abbildung 6.3: Modellergebnisse zum 1997 geschätzten Status (Y) der bis 24-jährigen Auszubildenden des Jahres 1996 - in Prozent

Um die Plausibilität dieser modellbasierten Befunde zusätzlich zu prüfen, werden in Abbildung 6.4 die Modellergebnisse für die Gruppe der Auszubildenden repliziert, die zum April 1996 eine Beschäftigungsdauer von mehr als zwei Jahren angegeben haben. Für diese überdurchschnittlich räumlich mobilen Personen wurden im NINR-Modell 20 besondere Ausfallzusammenhänge geschätzt.



Für die räumlich Immobilen zeigt der Vergleich der geschätzten und beobachteten Anteile (Zeilen 7-9 vs. Zeile 13 in Abb. 6.4) für das Modell 20 eine sehr gute Anpassung; die Abweichungen betragen maximal einen Prozentpunkt. Dies spricht insgesamt dafür, das NINR-Modell 20 zu akzeptieren. Die damit verbundenen Annahmen zum Ausfallgeschehen besagen, dass ein Abbruch der Ausbildung mit einer höheren räumlichen Mobilität einhergeht als der erfolgreiche Abschluss und dabei insbesondere Personen mit einer mehr als zweijährigen Ausbildungsdauer ein überdurchschnittliches Ausfallrisiko besitzen.

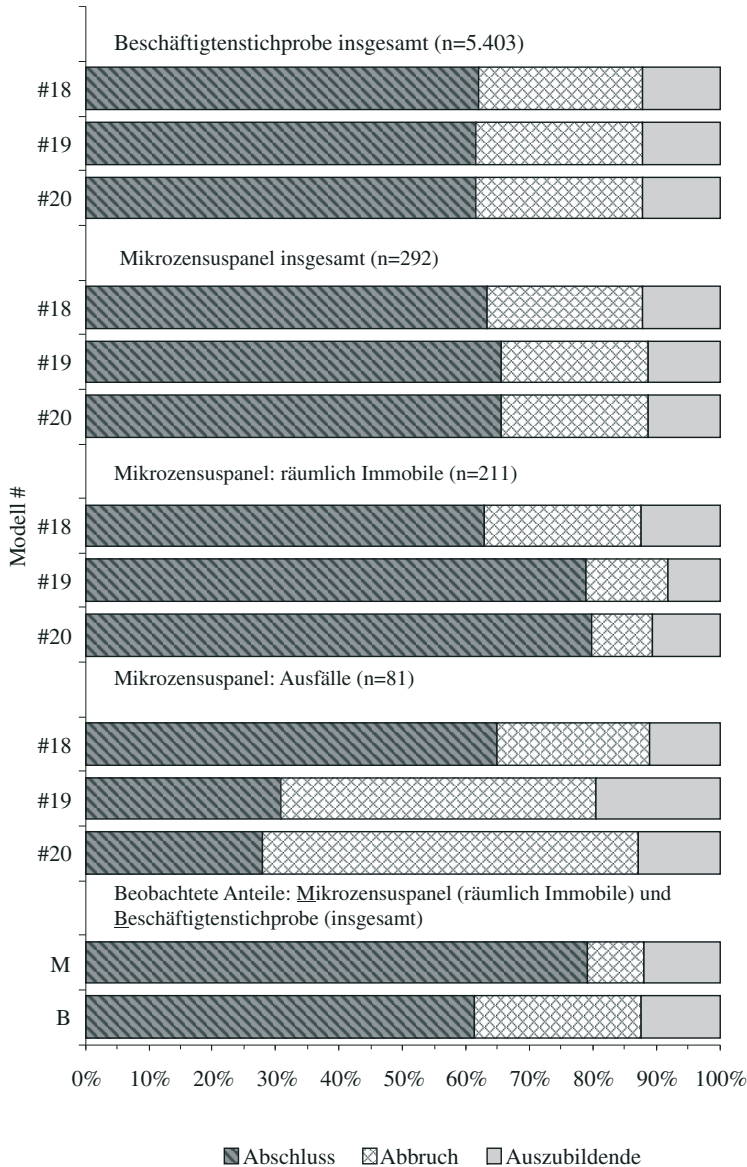


Abbildung 6.4: Geschätzter Ausbildungsstatus im April 1997 (Y) der bis 24-jährigen Auszubildenden des Jahres 1996 mit einer mehr als zweijährigen Beschäftigungsdauer - in Prozent

### 6.3 Können Gewichte die Stichprobenselektivität korrigieren?

Der Befund nicht ignorierbarer Ausfälle ist für Analysen zum Verlauf und Abschluss einer beruflichen Ausbildung im dualen System von weitreichenden Konsequenzen. Da die Annahme bedingt zufälligen Ausfalls (MAR) verworfen werden muss, kann im Grunde durch eine Gewichtung nur mit dem Kehrwert der Wahrscheinlichkeit für die Immobilität keine Ausfallkorrektur erreicht werden, sondern der Zusammenhang zwischen dem Ausfall und der abhängigen Variablen muss wie in dem obigen NINR-Modell explizit berücksichtigt werden.

Dies erfordert allerdings einigen Aufwand. Oft wird angenommen, dass in Analysen, in denen die MAR-Annahme berücksichtigt wird, der durch Ausfälle entstehende Bias ausreichend korrigiert werden kann: „In many realistic applications, however, we believe that departures from MAR are not large enough to effectively invalidate the results of an MAR-based analysis“ (Schaffer und Graham 2002: 164; ähnlich argumentieren Little und Rubin 2002: 346). Aus pragmatischer Sicht stellt sich deshalb die Frage, in welchem Maße mit Verzerrungen zu rechnen ist, falls auf der MAR-Annahme basierte Gewichte verwendet werden, obwohl diese nicht zutrifft. Diese Frage wird im Folgenden für Populationsschätzungen und statistische Modelle untersucht.

Schätzt man für räumlich immobile Auszubildende des Jahres 1996 den bis 1997 ermittelbaren Status nur unter Verwendung des Querschnittsfaktors ( $g_1$ ), zeigt Tabelle 6.5 Anteile, die sehr nahe an denen der hier zur „Eichung“ genutzten Beschäftigtenstichprobe liegen. Jedoch liegt der mit dem Querschnittsfaktor  $g_1$  ermittelte Gesamtwert von 806.200 Personen um 30 Prozent unter dem Gesamtwert der Beschäftigtenstichprobe ( $n = 1.152.200$ ). Zu beachten ist hierbei allerdings, dass mit dem Mikrozensuspanel bereits die Auszubildenden des Jahres 1996 um 14 Prozent im Vergleich zur Beschäftigtenstichprobe untererfasst sind (siehe Tab. 6.1).

Verwendet man zusätzlich zum Querschnittsfaktor eine Ausfallgewichtung mit dem Kehrwert der Wahrscheinlichkeit für räumliche Immobilität ( $g_2$ ), die der MAR-Annahme des obigen Modells 18 entspricht, kann die Untererfassung zwar auf 13,6 Prozent reduziert werden, jedoch weichen die Anteilswerte stärker von der Verteilung der Beschäftigtenstichprobe ab als bei reiner Querschnittsgewichtung. Insbesondere wird der Anteil der Absolventen überschätzt.

Ein ähnliches Bild zeichnet sich ab, wenn mit den im Teilprojekt „Methoden“ entwickelten Hochrechnungsfaktoren (Basic et al. 2005; Rohloff 2005) gewichtet wird. Sie setzen für eine wirksame Biaskorrektur voraus, dass innerhalb der Anpassungsgruppen kein Zusammenhang zwischen dem interessierenden Merkmal und dem Ausfall besteht (siehe Abschnitt 2.1.2, Seite 21). Diese auch als „response homogeneous group“ (Särndal et al. 1997: 577ff.) bezeichnete MAR-Annahme trifft hier aber nicht zu. In Bezug auf den Gesamtwert schneidet der Längsschnittfaktor, mit dem auch die räumliche Immobilität kontrolliert wird ( $g_4$ ), etwas schlechter ab als der einfache Längsschnittfaktor ( $g_3$ ).

Wie aufgrund der obigen Modelle nicht anders zu erwarten, kann durch eine Gewichtung mit dem Querschnittsfaktor und dem Kehrwert der mit dem NINR-Modell 20 geschätzten, vom Ausbildungsstatus abhängigen Antwortwahrscheinlichkeit ( $g_5$ ; siehe zur statistischen Begründung S. 35) die beste Anpassung an die Statusverteilung der Beschäftigtenstichprobe erreicht werden. Beim Gesamtwert bleibt jedoch die Untererfassung um rund 14 Prozent bestehen.

Tabelle 6.5: Statusverteilungen im Jahre 1997 der räumlich immobilten Auszubildenden im Mikrozensuspanel des Jahres 1996 bei verschiedenen Korrekturgewichtungen

Datenbasis Gewichtung (g)	darunter (in Prozent)				Relative Differenz*
	Insgesamt (in 1.000)	Auszu- bildender	Abschluss	Abbruch	
Beschäftigtenstichprobe g = 0,5 [2%-Stichprobe]	1.152,2	64,3	22,1	13,6	
Mikrozensuspanel					
Querschnittsfaktor 1997 g1 = BF750ki * BF750	806,2	63,4	25,4	11,2	-30,0 %
Querschnittsf. + MAR-Gew. g2 = g1/(1-P(Ausfall [M18]))	995,6	61,3	27,6	11,1	-13,6 %
Längsschnittfaktor g3 = gew_9697	1.001,6	61,7	27,2	11,1	-13,1 %
Längsschnittfakt. + Mobilität g4 = gewl_9697	975,0	62,3	26,9	10,8	-15,4 %
Querschnittsf. + NINR-Gew. g5 = g1/(1-P(Ausfall [M20]))	992,4	63,5	22,8	13,7	-13,9 %
Ungewichtete Stichprobe (n = 1.158)		63,9	25,5	10,6	

\*) Die relative Differenz bezieht sich auf den Gesamtwert der Beschäftigtenstichprobe.

*Gewichtungsvariablen:*

BF750ki: Korrekturfaktor zur Anpassung des Mikrozensuspanels an den Gesamtwert der Personen aus Privathaushalten im Erhebungsjahr 1997

BF750: Personen-Hochrechnungsfaktor für das Erhebungsjahr 1997 (Anpassung der Mikrozensus-Fallzahlen an Ergebnisse der laufenden Bevölkerungsfortschreibung mit Korrektur für den Ausfall nicht befragter Haushalte)

P(Ausfall [M18]): Durch ein Logit-Modell analog zum MAR-Modell 18 in Tabelle 6.4 geschätzte Wahrscheinlichkeiten für räumliche Mobilität

gew\_9697: Hochrechnungsfaktor für den Längsschnitt 1996-1997 ohne Berücksichtigung von Umzügen

gewl\_9697: Hochrechnungsfaktor für den Längsschnitt 1996-1997 mit Schätzung der räumlichen Immobilität zur Berücksichtigung von Umzügen

P(Ausfall [M20]): Durch das NINR-Modell 20 in Tabelle 6.4 geschätzte Ausfallwahrscheinlichkeiten

*Quelle:* Mikrozensuspanel 1996-1999, Beschäftigtenstichprobe (siehe Abbildung 6.2 und Tabelle A.6); eigene Berechnungen.

Abbildung 6.5 zeigt für weitere Merkmale, wie unterschiedlich die Zusammenhänge mit dem bis 1997 erreichten bzw. ermittelten Status abgebildet werden, wenn lediglich die räumlich Immobilen des Mikrozensuspanels analysiert werden. Da sich durch die verschiedenen Gewichtungsvariablen (g2-g4) keine gravierenden Unterschiede der Anteile ergeben, werden der Übersichtlichkeit halber nur die Ergebnisse der Querschnittsgewichtung mit Logit-Gewicht (g2) berichtet. Daneben sind, äquivalent zu einer Gewichtung mit dem NINR-Hochrechnungsfaktor (g5), die Verteilungen der Beschäftigtenstichprobe und Schätzungen durch das Modell 20 (M20; ungewichtete Daten) abgebildet.

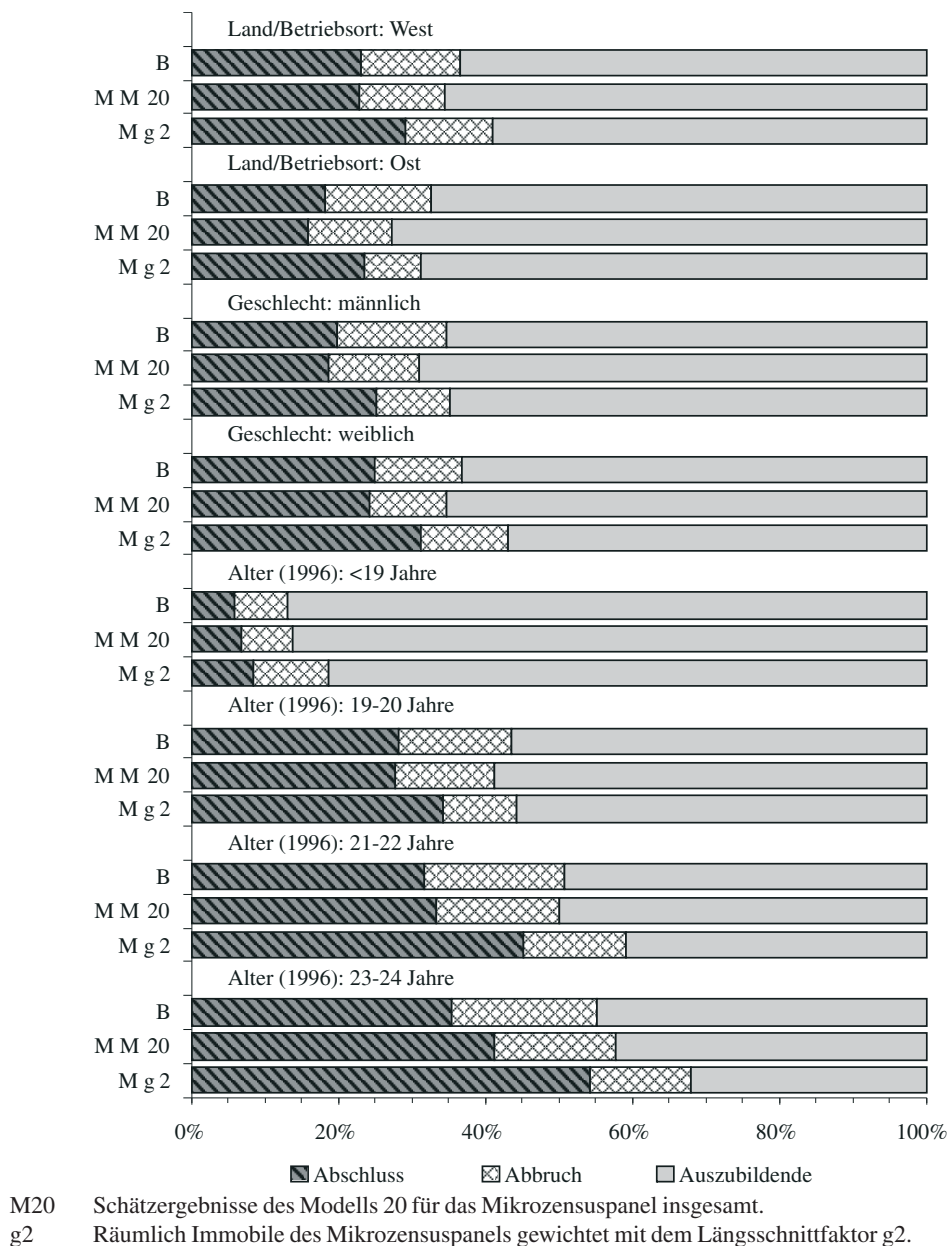


Abbildung 6.5: Anteile des Status 1997 in der Beschäftigtenstichprobe (B) insgesamt und für räumlich immobile Auszubildende des Jahres 1996 im Mikrozensuspanel (M) für ausgewählte Merkmale bei verschiedenen Modellannahmen – in Prozent.

Unter der nicht zutreffenden Annahme bedingt zufälliger Ausfälle werden mit der MAR-Ausfallgewichtung (g2) teilweise deutlich andere Chancenverhältnisse geschätzt. Beispielsweise betragen für Männer (Frauen) die Prozentsatzverhältnisse der Anteile für Abschluss vs. Abbruch in der Beschäftigtenstichprobe 1,3 (Frauen: 2,1), beim Modell 20 des Mikrozensuspanel 1,5 (Frauen:

en: 2,3), bei Ausfallgewichtung dagegen 2,5 (Frauen: 2,6). Stellt man den Chancenverhältnissen der Frauen die der Männer gegenüber, liegen die entsprechenden Odds-Ratios für die Beschäftigtenstichprobe ( $2,1 / 1,3 = 1,6$ ) und das Modell 20 (1,5) nahe beieinander. Die Odds-Ratios zeigen, dass Frauen größere Chancen eines erfolgreichen Ausbildungsabschlusses besitzen als Männer. Dieser Chancenvorteil von Frauen kommt dagegen nicht zum Ausdruck, wenn für die gewichteten (g2) Daten räumlich immobiler Personen das Odds-Ratio berechnet wird ( $2,6 / 2,5 = 1$ ).

Diese Befunde entsprechen den Ergebnissen der log-linearen Modelle, die zugleich darauf hinweisen, dass statistische Modelle, die unter der Annahme bedingt zufälliger Ausfälle (MAR) im Mikrozensuspanel geschätzt werden, zu anderen Ergebnissen als bei der Beschäftigtenstichprobe führen. Wenn ergänzend zu den log-linearen Modellen die Konsequenzen ausfallgewichteter Daten räumlich mobiler Auszubildender analysiert werden, kommt für den Vergleich zwischen Mikrozensuspanel und der Beschäftigtenstichprobe der Wald-Test in Frage.

Die Gewichtungsvariablen des Mikrozensuspanels wurden auf die beobachtete Fallzahl normiert. Die Analysen der Beschäftigtenstichprobe sind ungewichtet. Tabelle A.5 im Anhang zeigt, dass die Ergebnisse von Logit-Modellen zum Statusübergang für räumlich immobile Personen des gewichteten Mikrozensuspanels von den Ergebnissen der Beschäftigtenstichprobe signifikant abweichen. Dies trifft sowohl für ungewichtete als auch die verschiedenen gewichteten Daten zu, wobei der Längsschnittfaktor + Mobilität (g4) den geringsten  $\chi^2$ -Wert aufweist. Betrachtet man die einzelnen Tests auf Gleichheit der Koeffizienten der Beschäftigtenstichprobe und des Mikrozensuspanels (siehe Spalte  $P_w$  in Tabelle A.6),<sup>72</sup> zeigen sich durchgängig Unterschätzungen für die Koeffizienten zum Ausbildungsabbruch insgesamt (Konstante) und für Abbrecher mit einer Beschäftigungsdauer über zwei Jahren. Bei den räumlich immobilen Absolventen weichen die Koeffizienten zum allgemeinen Bildungsabschluss und zu den Bauberufen von den Schätzungen auf Basis der Beschäftigtenstichprobe ab. Diese Befunde entsprechen der obigen deskriptiven Analyse. Zusammenfassend ist somit festzuhalten, dass mit den verschiedenen MAR-basierten Gewichten bei den hier vorliegenden nicht ignorierbaren Ausfällen keine wirksame Biasreduktion erzielt werden kann.

Kombiniert man jedoch die MAR-basierten Gewichte provisorisch mit den durch das Selektionsmodell unter der Annahme nicht ignorierbarer Ausfälle geschätzten Antwortwahrscheinlichkeiten, ergeben sich im Vergleich zu den Referenzdaten deutlich bessere Schätzungen. Für andere praktische Anwendungen deutet dies zugleich auf Möglichkeiten der Biaskorrektur hin. Würden die soziodemografischen Hilfsvariablen vorliegen, die bei Konstruktion der Hochrechnungsfaktoren bzw. Gewichte verwendet wurden, könnten die modellspezifischen Antwortwahrscheinlichkeiten bei der Kalibrierung als Eingangsgewichte eingesetzt werden.

## 6.4 Chancen der Beschäftigung im erlernten Beruf

Auf Basis von Querschnittsdaten des Mikrozensus waren bisher nur eingeschränkte Auswertungen zum Einstieg ins Erwerbsleben von Berufsanfängern möglich (Brauns et al. 2000; Handl 1996). Das Hauptproblem bestand in der Abgrenzung der Zielpopulation der Berufsanfänger, die nur mit Hilfe von nicht prüfbar Annahmen über das Ende einer Ausbildung bzw. das Eintrittsdatum ins Beschäftigungssystem erfasst werden können. Da seit 1996 in der EU-Arbeitskräftestichprobe in vierjährigem Turnus das Datum des beruflichen Abschlusses erfragt wird, liegen die-

72 Bei diesen Werten wurde nicht für multiples Testen kontrolliert. Verwendet man die Bonferroni-Korrektur, sind alle P-Werte nicht mehr statistisch signifikant. Allerdings wird die Bonferroni-Korrektur in der Praxis oft als zu konservativ eingeschätzt.

se Informationen für die Substichprobe vor. Allerdings wird nach wie vor der erlernte Beruf nicht im Mikrozensus erfragt.<sup>73</sup> Mit dem Mikrozensuspanel sind nun solche Themen bearbeitbar. Das vorwiegende Ziel dieses Abschnitts ist es, methodische Fragen zu diskutieren, welche die mit dem Panel erstmals durchführbaren Vergleiche zwischen erlerntem und ausgeübtem Beruf betreffen.

Für die Untersuchung der Chancen einer Beschäftigung im erlernten Beruf nach Abschluss der Ausbildung stellt aber das Ergebnis nicht ignorierbarer Ausfälle vor allem deshalb ein Problem dar, weil die Analysen im Grunde genommen von einem zweistufigen Selektionsprozess ausgehen müssen. Auf der ersten Stufe sind die durch die räumliche Mobilität zustande gekommenen selektiven Ausfälle zu berücksichtigen bevor im zweiten Schritt die eigentlich interessierende Frage bearbeitet werden kann, mit welchen Selektionsprozessen die Beschäftigung im Ausbildungsberuf verbunden sind.

Verwendet man wieder die Beschäftigtenstichprobe als Referenzstatistik, sind zudem Übergänge in Zustände ohne Meldung in der Beschäftigtenstichprobe („Meldelücken“) zu beachten und analog beim Mikrozensuspanel abzugrenzen. Wegen des resultierenden geringen Stichprobenumfangs räumlich immobiler Ausbildungsabsolventen beim Mikrozensuspanel in den einzelnen Teilgruppen (siehe Tabelle 6.6) werden im Folgenden nur deskriptive und explorative Analysen angestrebt. Diese aus pragmatischen Gründen notwendige Beschränkung auf räumlich immobile Absolventen hat einerseits entsprechend höhere Stichprobenfehler zur Folge, andererseits können dabei selektive Ausfälle nicht ausgeschlossen bzw. effektiv kontrolliert werden. Allerdings deuten die Abweichungen ungewichteter Daten der Absolventen des Mikrozensuspanels im Vergleich zur Beschäftigtenstichprobe (siehe Tabelle A.7) und die Ergebnisse der obigen log-linearen Modelle darauf hin, dass für diese Gruppe nur geringe Selektivitätseffekte vorliegen. Zunächst wird ein knapper Überblick zu den theoretischen Konzepten und zum Forschungsstand gegeben und die Konstruktion der Variablen zum Berufswechsel beschrieben.

#### **6.4.1 Theoretische Konzepte und Forschungsstand im Überblick**

In Bezug auf die Verbindung zwischen dem Ausbildungs- und Berufssystem wird in Deutschland häufig das sogenannte Schwellenkonzept des IAB zur Orientierung verwendet. Mit der ersten Schwelle werden Übergänge vom allgemein bildenden Schulsystem in eine berufliche Erstausbildung thematisiert. Die daran anschließende Allokation von Personen mit einer abgeschlossenen Berufsausbildung auf berufliche Positionen im Beschäftigungssystem stellt die hier zentrale zweite Schwelle dar.

Die Berufsausbildungen im dualen System zeichnen sich insbesondere im internationalen Vergleich dadurch aus, dass die erreichten Abschlüsse infolge der Standardisierung einerseits allgemein anerkannt sind und von mehreren Unternehmen nachgefragt werden (Müller 2005). Andererseits sind sie auf bestimmte Berufe beschränkt, somit vorwiegend auf Teilarbeitsmärkten einsetzbar und die Auszubildenden erwerben in den Betrieben betriebsspezifische Kenntnisse.

Diese allgemeinen und betriebsspezifischen Qualifikationen werden auch in der Humankapitaltheorie (vgl. Franz 1999: 81ff.) unterschieden. Gemäß der Humankapitaltheorie sind Auszubildende nach dem Abschluss bestrebt, ihr Einkommen zu maximieren. Längerfristig ist dabei häufig die weitere Beschäftigung im erlernten Beruf vorteilhaft. Die Ausbildungsbetriebe sind im Allgemeinen daran interessiert, die von den Auszubildenden erworbenen allgemeinen und insbesondere die betriebsspezifischen Fertigkeiten weiter zu nutzen. Im Vergleich zur Einstellung betriebsfremder Fachkräfte können die Betriebe während der Ausbildung die Leistungsfähigkeit ih-

73 Seit 2005 wird zwar die Fachrichtung des höchsten beruflichen Ausbildungsabschlusses erfragt, jedoch erlauben die Antwortkategorien nur eine grobe Klassifikation von Ausbildungsberufen.

rer Auszubildenden und künftigen Mitarbeiter in der Regel sehr gut einschätzen und sparen somit Einstellungs- und Einarbeitungskosten. Dieser Aspekt spielt auch in suchtheoretischen Ansätzen eine wichtige Rolle.

Das betriebliche Interesse an qualifizierten Arbeitskräften mit guten betriebspezifischen Kenntnissen wird besonders von segmentationstheoretischen Ansätzen betont. Für Deutschland wird mit besonderem Bezug auf Qualifikationsanforderungen von einem dreigeteilten Arbeitsmarkt ausgegangen (Sengenberger 1987): Arbeitsmärkte mit unspezifischen Jedermannsqualifikationen, berufsfachliche und interne bzw. betriebliche Arbeitsmärkte. In Übereinstimmung mit den Argumenten der Humankapitaltheorie ist davon auszugehen, dass Auszubildende aufgrund ihrer erworbenen betriebspezifischen Qualifikationen von den Ausbildungsbetrieben gute Weiterbeschäftigungsangebote erhalten. Jedoch endet mit der Ausbildung auch der Arbeitsvertrag mit dem Ausbildungsbetrieb. Bei konjunkturellen Engpässen auf dem Arbeitsmarkt ist deshalb diese Gruppe, insbesondere im Unterschied zu Kernbelegschaften, mit besonderen Schwierigkeiten konfrontiert. In einigen Wirtschaftsbereichen konnten die Gewerkschaften mit den Arbeitgeberverbänden tarifliche Regelungen zur Übernahme der Auszubildenden vereinbaren (Bispingk et al 2002).

Nach Analysen von Steinmann (2000: 158ff., 255f.) sind betriebliche Merkmale, insbesondere die Betriebsgröße (vgl. auch BMBF 2001: 475 [S. 12/Anlage 2]), in Bezug auf eine ausbildungsadäquate Beschäftigung wesentlich wichtiger als individuelle Merkmale der Auszubildenden. Konietzka (2004: 292) berichtet zusammenfassend für Ausländer geringere Chancen der Beschäftigung im erlernten Beruf als für Deutsche. Steinmann (2000) und Konietzka (2004) finden im Zeitvergleich bis Anfang der 1990er Jahre eine rückläufige Tendenz des Verbleibs im Ausbildungsbetrieb. Nach Ergebnissen von Haas (2002) ist dagegen auf der Basis der Beschäftigtenstichprobe zwischen 1992 und 1997 wieder eine Zunahme zu beobachten, d. h. eine abnehmende Tendenz von Betriebswechseln nach dem Ende der Ausbildung. Sowohl in West- als auch Ostdeutschland ist ein Rückgang der Berufswechsler<sup>74</sup> festzustellen, wobei sich die Anteile in den neuen Bundesländern von 32 Prozent im Jahre 1992 auf 22 Prozent im Jahre 1997 dem Niveau der alten Bundesländer (1992: 25 %; 1997: 20,6 %) angenähert haben. Von den erfolgreichen Ausbildungsabsolventen des Jahres 1995 arbeiteten nach einem halben Jahr nur noch 85 bzw. 81 Prozent in ihrem in West- bzw. Ostdeutschland ansässigen Betrieb (Haas 2002: 20).

## **6.4.2 Abgrenzung der Stichprobe und Variablenkonstruktion**

Weil Betriebswechsel mit Berufswechseln und Umzügen verbunden sein können, ist dieser Befund auf Basis der Beschäftigtenhistorik auch für die Analyse des Mikrozensuspanels zu den Chancen einer Beschäftigung im erlernten Beruf von Bedeutung. Im Mikrozensuspanel stehen allerdings für diese Analysen nur noch die räumlich immobil Absolventen zur Verfügung, in der Beschäftigtenstichprobe nur Absolventen mit Meldungen zu einer sozialversicherungspflichtigen Beschäftigung oder zum Leistungsempfang. Um das Mikrozensuspanel mit der Beschäftigtenstichprobe vergleichbar zu machen, sind weitere Selektionen notwendig. Der Ausschluss von Absolventen, für die zum Zeitpunkt nach dem Ausbildungsabschluss keine Meldungen in der Beschäftigtenstichprobe vorliegen, betrifft im Mikrozensuspanel rund 15 Prozent der kumulierten

74 Dabei ist zu beachten, dass die Abgrenzung anerkannter Ausbildungsberufe durch Verwendung der Berufsordnungen (Dreisteller) der Berufsklassifikation mit Unschärfen verbunden ist. Zudem werden mit der Beschäftigtenstatistik Berufswechsel eher unterschätzt, da die Berufsangaben nicht zu den meldepflichtigen Merkmalen gehören (Haas 2002). Die Berufswechsel wurden auf der Basis von Berufsordnungen identifiziert.



Übergänge 1996/97, 1997/98 und 1998/99 (s. Tabelle 6.6). Daneben ist zu beachten, dass die Gegenüberstellung des Ausbildungsberufes und des nach Abschluss der Ausbildung ausgeübten Berufes nur für Erwerbstätige möglich ist.<sup>75</sup>

Durch diese Selektion werden Teilfragen zur Verwertbarkeit des Berufsabschlusses auf dem Arbeitsmarkt ausgeblendet. Der Ausschluss von Erwerbslosen im Mikrozensuspanel betrifft immerhin rund 13 Prozent der Absolventen, sodass schließlich nur noch 73 Prozent von insgesamt 1.011 erfolgreichen und räumlich immobilten Absolventen der drei Übergänge für nachfolgende Analysen zur Verfügung stehen. Dagegen können dafür aus der Beschäftigtenstichprobe 85 Prozent der Absolventen herangezogen werden.

*Tabelle 6.6:* Abgrenzung der Stichproben für die Analysen zum Übergang von Ausbildungsabsolventen in eine anschließende Beschäftigung

Population (Übergänge 1996/97, 1997/98 und 1998/99)	Mikrozensuspanel		Beschäftigtenstichprobe	
	n	in %	n	in %
15- bis 24-jährige Auszubildende in anerkannten Ausbildungsberufen zum Zeitpunkt $t_0$ (1996, 1997, 1998) mit Erstausbildung und Ausbildungsabschluss bis zum Zeitpunkt $t_1$ (1997, 1998, 1999; Mikrozensuspanel: nur räumlich Immobiler)	1.011	100	15.572	100
- Ohne sozialversicherungspflichtige Beschäftigungsmeldungen in der Beschäftigtenstichprobe („Melde-lücke“ $t_1$ ); Mikrozensuspanel: Ausschließlich geringfügig Beschäftigte, Selbstständige, mithelfende Familienangehörige, Beamte, Soldaten, Wehrdienstleistende, Zivildienstleistende, Nichterwerbspersonen	149	14,7	1.323	8,5
= Personen mit Meldungen in der Beschäftigtenstichprobe zum Zeitpunkt $t_1$	604	59,7	14.249	91,5
- Meldungen aus Leistungsempfängerdatei (Beschäftigtenstichprobe) bzw. Erwerbslose (Mikrozensuspanel) $t_1$	128	12,7	970	6,2
= Sozialversicherungspflichtig Beschäftigte zum Zeitpunkt $t_1$ (1997, 1998, 1999)	734	72,6	13.279	85,3

*Quelle:* Mikrozensuspanel 1996-1999; IAB-Beschäftigtenstichprobe (IABS-R01); ungewichtete Fallzahlen; eigene Berechnungen.

Nach Ergebnissen der hier verwendeten Daten der Beschäftigtenstichprobe haben 60 Prozent der Absolventen die Ausbildung im Sommer 1996 abgeschlossen. Rund 24 Prozent beendeten die Ausbildung zum Jahresende. Damit bezieht sich die im Folgenden betrachtete Situation des regulären Berufseinstiegs auf eine Zeit von durchschnittlich neun Monaten (Median) nach Abschluss der Berufsausbildung.

Berufswechsel nach der Ausbildung können ganz unterschiedlich thematisiert werden; beispielsweise als Indikator für Abstimmungsprobleme zwischen Bildungs- und Beschäftigungssystem. Eine Reihe von Analysen zum Übergang nach der Ausbildung im dualen System behandeln Fragen zu den Chancen der Beschäftigung als Fachkraft oder den Risiken unterwertiger Be-

<sup>75</sup> Für Westdeutschland können zwar ab April 1999 auch geringfügig Beschäftigte unterschieden werden, aus Gründen der Vergleichbarkeit mit den Übergängen 1996/97 und 1997/98 wird aber auf diese Differenzierung verzichtet.



schäftigung und der Statuskongruenz (siehe dazu u. a. Büchel und Weißhuhn 1995; Konietzka 1999; Wolbers 2003). Beim folgenden Vergleich des Ausbildungsberufes mit dem ausgeübten Beruf geht es aber nicht um solche Aspekte vertikaler beruflicher Mobilität, sondern darum, etwas darüber zu erfahren, ob die Absolventen im erlernten oder zumindest in einem verwandten Beruf tätig oder ob sie ausbildungsfern beschäftigt sind.

Je differenzierter die hierfür verwendete Berufsklassifikation ist, umso mehr Berufswechsel wird man dabei feststellen. Beispielsweise werden für die Ausbildungsabsolventen der Übergänge 1996/97 bis 1998/99 im Mikrozensuspanel auf der Basis von Berufsordnungen 33 Prozent, auf der Grundlage von Berufsgruppen jedoch nur 23 Prozent Berufswechsel festgestellt (siehe Tabelle 6.7). In ähnlicher Weise wechseln in der Beschäftigtenstichprobe rund 22 Prozent der Auszubildenden nach dem Abschluss den Ausbildungsberuf, wenn die in den Daten enthaltenen 130 Berufskategorien zugrunde gelegt werden, jedoch nur 15 Prozent, wenn die recodierten Berufsabschnitte (25 Kategorien) verwendet werden.

*Tabelle 6.7:* Anteile von Berufswechseln bei Ausbildungsabsolventen der Übergänge 1996/97, 1997/98 und 1998/99 in Abhängigkeit von verschiedenen differenzierten Berufsangaben (in Prozent)

Berufsklassifikation	Mikrozensus- panel	Beschäftigten- stichprobe
BA-Berufsabschnitte (25 Kategorien)	17,4	14,7
KldB 1988 BA Berufsabschnitte/-gruppen (IABS-R01-Code)		21,5
KldB 1992 StBA recodiert in IABS-R01-Code	26,0	
KldB 1992 StBA Berufsordnungen (Dreisteller)	33,1	
KldB 1992 StBA Berufsgruppen (Zweisteller)	22,5	
Retrospektivfrage EU-Arbeitskräftestichprobe	11,2	

*Quelle:* Mikrozensuspanel 1996-1999 (n = 734); Beschäftigtenstichprobe (IABS-R01; n = 13.279); eigene Berechnungen.

Diese Unterschiede liegen einerseits daran, dass die einzelnen Ausbildungsberufe Schlüsselqualifikationen vermitteln, die eine qualifizierte Beschäftigung auch in benachbarten Berufen gewährleisten. Andererseits sind differenziertere Berufsangaben deutlich stärker mit Codierungsfehlern behaftet (Elias 1997), sodass zeitlich inkonsistente Angaben, die beim Vergleich tief gegliederter Berufsangaben als Veränderung gezählt werden, zu höheren Wechselanteilen führen.

Um diese Probleme zu minimieren und um Wechsel zwischen benachbarten Berufsfeldern als beruflich adäquate Tätigkeiten anzuerkennen, erscheint es deshalb sinnvoll, die Analyse auf Basis des Vergleichs der Berufsabschnitte (25 Berufskategorien) vorzunehmen. Hierbei können Unterschiede und Ähnlichkeiten zwischen Berufen nur näherungsweise berücksichtigt werden; nur Wechsel zwischen Berufen verschiedener Berufsabschnitte werden als Wechsel gezählt (vgl. Fitzenberger und Spitz 2004; Schaeper et al. 2000). In diesem Fall wird für beide Datenquellen ein ähnlicher Berufswechsel festgestellt, wobei der Anteil im Mikrozensuspanel mit 17 Prozent um rund drei Prozentpunkte höher als in der Beschäftigtenstichprobe liegt. Die Differenz kann u. a. mit der Selektion räumlich immobiler Absolventen, aber auch mit der im Vergleich zur Beschäftigtenstichprobe unterschiedlichen Berufscodierung (Haushaltsbefragung mit Proxy-Interviews vs. Angaben durch die Betriebe) zusammenhängen.

Zu beachten ist, dass in der Beschäftigtenstichprobe Berufswechsel tendenziell untererfasst sind. Da die Berufsangaben nicht zu den meldepflichtigen Merkmalen gehören und durch die Be-

triebe i. d. R. fortgeschrieben werden, bleiben innerbetriebliche Berufswechsel oft unerkannt (Haas 2002). Andererseits können Berufswechsel, die im Zusammenhang mit Betriebswechseln stattfinden, übererfasst sein, da in diesem Fall die Auskunft gebende Stelle wechselt.

Im Mikrozensus bietet sich für eine einfache Abschätzung der Qualität des ermittelten Berufswechsels der Vergleich mit den retrospektiv erfragten Angaben zum Berufswechsel an. Die Frage „Haben Sie seit Ende ... [Vorjahr] den ausgeübten Beruf gewechselt?“ wird jedoch nur von elf Prozent bejaht. Der Vergleich zwischen den Retrospektivangaben und den auf Basis der Berufsabschnitte ermittelten Werten zeigt lediglich eine Übereinstimmung von 82 Prozent. Die Übereinstimmung sinkt auf 71 Prozent, wenn die Berufswechsel durch den Vergleich der Berufsordnungen (Dreisteller) berechnet werden.

Es liegt nahe zu vermuten, dass die bei den Retrospektivangaben, auch im Vergleich zur Beschäftigtenstichprobe, zu beobachtende Unterschätzung des Berufswechsels nicht nur mit den bei diesen Fragen häufig entstehenden Rückerinnerungsproblemen, sondern auch mit Proxy-Interviews zusammenhängt, da in diesem Fall die Auskunftgebenden nicht immer in ausreichendem Maße über detaillierte Kenntnisse verfügen. Dies lässt sich mit den verfügbaren Daten allerdings nicht untersuchen. Ersatzweise kann ein anderer Indikator zur Antwortkonsistenz herangezogen werden. Im Rahmen der Zusammenführung der Querschnittsdaten wurden zeitlich inkonsistente Angaben zum Geburtsjahr und Geschlecht nach Dateninspektion bereinigt. Differenziert man danach, weisen die rund zehn Prozent der Ausbildungsabsolventen mit nachträglichen Korrekturen einen Berufswechselanteil von 19,5 Prozent auf, der um 2,3 Prozentpunkte höher ist als in der Gruppe mit zeitlich konsistenten Angaben zum Geburtsjahr und Geschlecht (hier ohne Nachweis).

Lässt man die Stichprobenselektivität außer Acht kann zusammenfassend zur Differenz der berechneten Anteile des Berufswechsels zwischen Mikrozensuspanel und Beschäftigtenstichprobe festgehalten werden, dass im Mikrozensuspanel zeitlich inkonsistente Angaben zu einer tendenziellen Überschätzung des Berufswechsels führen. Dagegen stellen die Ergebnisse der Beschäftigtenstichprobe eher eine Untergrenze dar, da innerbetriebliche Berufswechsel untererfasst sein dürften. Die Unterschiede zwischen beiden Datensätzen erscheinen insgesamt jedoch nicht so gravierend, dass sie gegen einen Vergleich sprechen würden. Gleichwohl müssen diese Punkte bei der Interpretation der Ergebnisse beachtet werden.

### 6.4.3 Deskriptive Analysen zum Berufswechsel

Die Übergänge vom Ausbildungs- zum ausgeübten Beruf können im Mikrozensuspanel nur für die räumlich immobilen Absolventen beobachtet werden. In Bezug auf die Ausfälle bzw. Wegzüge dieser Gruppe haben die obigen Analysen gezeigt, dass erfolgreiche Ausbildungsabsolventen weniger räumlich mobil sind als Ausbildungsabbrecher und Auszubildende (vgl. Abb. 6.2, Tab. A.5 und Seite 126). Einerseits lässt dies vermuten, dass die Anteile von Berufswechseln im Mikrozensuspanel und in der Beschäftigtenstichprobe weitgehend gleich sein werden; abgesehen von den oben angesprochenen grundlegenden Unterschieden zwischen den Daten. Andererseits sind beim Übergang 1996/97 insgesamt, d. h. einschließlich der Auszubildenden und Ausbildungsabbrecher, statistisch signifikante Ausfälle in den Berufsgruppen der Waren- und Dienstleistungskaufleute (Gruppe 4) und der Verkehrs- bzw. allgemeinen Dienstleistungsberufe (Gruppe 6) aufgetreten. Für diese Berufsgruppen ist deshalb mit ausfallbedingten Abweichungen zwischen Mikrozensuspanel und Beschäftigtenstichprobe zu rechnen. Geht man von einem Zusammenhang zwischen Berufswechsel und räumlicher Mobilität aus, d. h. dass fortziehende Perso-

nen häufiger als räumlich immobile Personen ihren Beruf wechseln, ist zu erwarten, dass das Ausmaß von Berufswechseln dieser Berufsgruppen im Mikrozensuspanel unterschätzt wird.

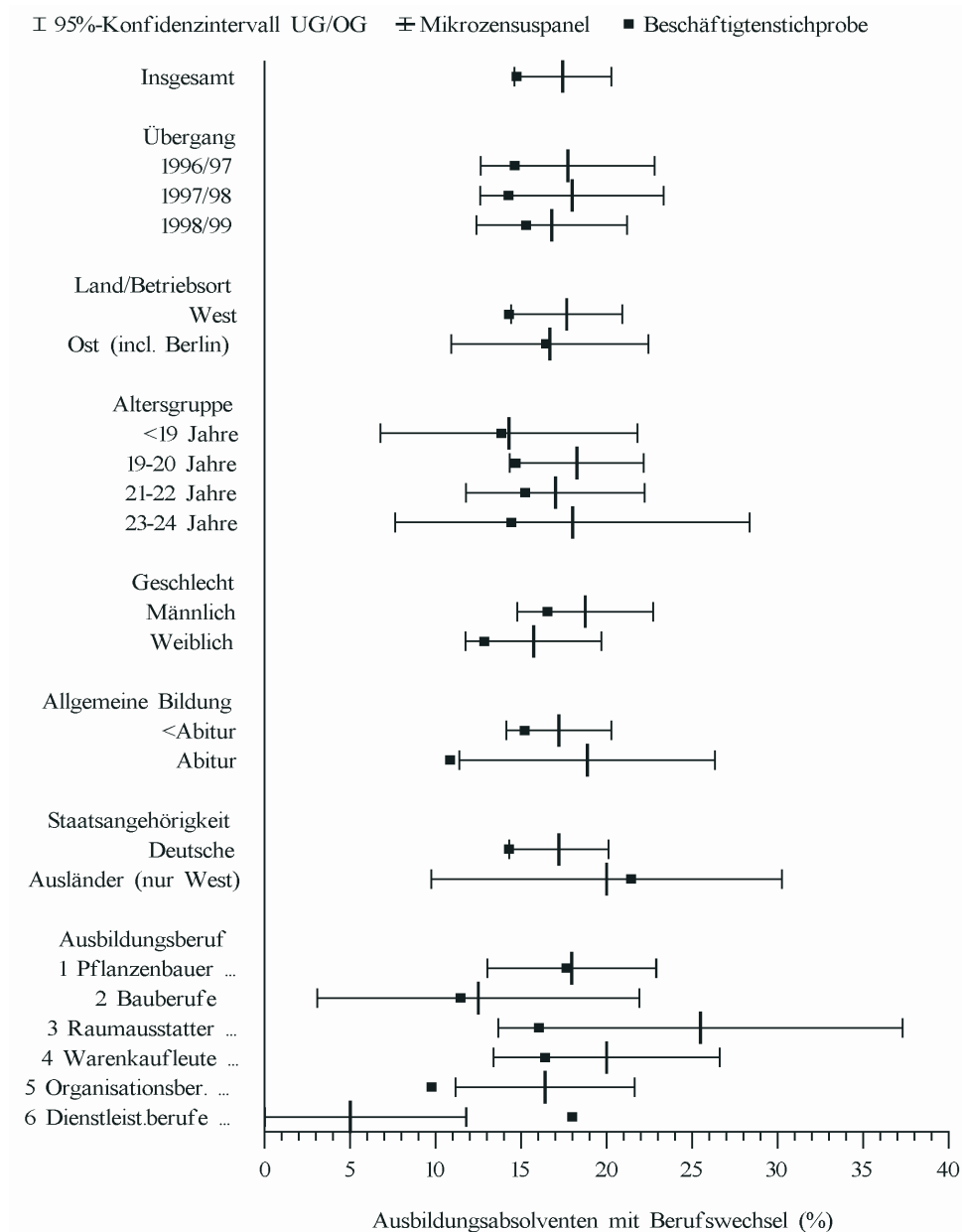
Abbildung 6.6 vergleicht die Anteile von Ausbildungsabsolventen der kumulierten Übergänge der Beschäftigtenstichprobe als Referenzstatistik mit den Anteilen des Mikrozensuspanels. Zusätzlich werden zu den Anteilen 95 %-Konfidenzintervalle<sup>76</sup> abgebildet (siehe auch Tabelle A.7). Die betrachteten Merkmale beziehen sich immer auf das Jahr vor dem Übergang in den Beruf nach Abschluss der Ausbildung. Abgesehen von der Gesamtdifferenz der Anteile des Berufswechsels zwischen beiden Datenbasen (Mikrozensuspanel 17,4 %; Beschäftigtenstichprobe 14,7 %) zeigen sich teilweise ähnliche, teilweise aber auch gegenläufige Zusammenhangsmuster. Beispielsweise wechseln weibliche Ausbildungsabsolventen sowohl in der Beschäftigtenstichprobe (12,8 %) als auch im Mikrozensuspanel (15,7 %) seltener als Männer (16,6 % bzw. 18,8 %) den Beruf. Obwohl weibliche Auszubildende im Mikrozensuspanel häufiger als Männer ausfallen, sind dennoch keine gravierenden Unterschiede zwischen beiden Datenquellen festzustellen, selbst wenn nur die räumlich immobilten Absolventen des Mikrozensuspanels betrachtet werden.

In der Beschäftigtenstichprobe haben Auszubildende, die im Vorjahr in ostdeutschen Betrieben arbeiteten, ein leicht höheres Risiko des Berufswechsels als Auszubildende westdeutscher Betriebe (16,5 % vs. 14,3 %). Gleichfalls wechseln in der Beschäftigtenstichprobe Personen ohne Abitur häufiger den Ausbildungsberuf als Personen mit Abitur (15,2 % vs. 10,9 %). Ein höheres allgemeines Bildungsniveau reduziert offensichtlich das Risiko eines Berufswechsels. Im Mikrozensuspanel werden dagegen für Auszubildende aus ostdeutschen Betrieben (16,7 % vs. West: 17,7 %) als auch für Auszubildende ohne Abitur (17,2 % vs. Abitur: 18,9 %) jeweils geringfügig niedrigere Anteile von Berufswechseln geschätzt als für die entsprechenden Vergleichsgruppen. Diese Unterschiede könnten bei Auszubildenden aus ostdeutschen Betrieben mit der überdurchschnittlichen Ausfallneigung verbunden sein (vgl. Abb. 6.2 und Tab. A.5). Dies trifft für den Unterschied hinsichtlich des allgemeinen Bildungsabschlusses aber vermutlich nicht zu, da dazu kein signifikanter Ausfalleffekt festgestellt werden konnte (siehe Tab. 6.4).

Berücksichtigt man beim Mikrozensuspanel die breiten Konfidenzintervalle bzw. den hohen Stichprobenfehler, sind diese Unterschiede allerdings kaum interpretationsfähig. In diesem Zusammenhang ist auch zu beachten, dass die Konfidenzintervalle des Mikrozensuspanels i. d. R. die mit der Beschäftigtenstichprobe berechneten Anteile abdecken.

Stärkere Differenzen sind bei zwei Berufsgruppen zu beobachten. Entgegen den Vermutungen aufgrund der Ausfallanalysen zeigen sich bei Waren- und Dienstleistungskaufleuten (Gruppe 4) keine gravierenden Differenzen zwischen Mikrozensuspanel und Beschäftigtenstichprobe. Statt dessen ist das aber bei den Organisations- und Erziehungsberufen (Gruppe 5) der Fall, für die mit dem Mikrozensuspanel mit 16 Prozent ein um rund sieben Prozentpunkte höherer Berufswechsel als mit der Beschäftigtenstichprobe geschätzt wird. Wie zu erwarten, finden sich erhebliche Unterschiede des Berufswechsels für allgemeine Dienstleistungs- und Verkehrsberufe (Gruppe 6). Während in der Beschäftigtenstichprobe 18 Prozent der Ausbildungsabsolventen dieser Gruppe den Beruf wechseln, ermittelt man mit dem Mikrozensuspanel lediglich fünf Prozent.

76 Bei der Berechnung des Stichprobenfehlers im Mikrozensuspanel wird die Klumpung berücksichtigt.



Die Merkmale beziehen sich auf den Zeitpunkt des Jahres vor dem Ausbildungsabschluss.

Quelle: Mikrozensuspanel 1996-1999; Beschäftigtenstichprobe (IABS-R01); siehe Tab. 6.7 und Tab. A.7.

Abbildung 6.6: Berufswechsel von Ausbildungsabsolventen im Mikrozensuspanel und in der Beschäftigtenstichprobe (in Prozent)

Aufgrund der systematischen Unterschiede der Datengenerierung (Umfrage- vs. prozessproduzierte Daten), aber vor allem auch aufgrund des Stichprobenfehlers sind diese Abweichungen nur eingeschränkt zu interpretieren. Schätzt man nämlich mit dem Mikrozensuspanel ein Logit-Modell zum Berufswechsel, ist im Unterschied zur Beschäftigtenstichprobe kein statistisch signifikanter Zusammenhang der in Abbildung 6.6 dargestellten Merkmale festzustellen. Zudem unterscheiden sich die Regressionskoeffizienten nicht signifikant zwischen den beiden Datensätzen (siehe Tabelle A.7).

Zusammenfassend ist daher zu den Chancen der Beschäftigung im erlernten Beruf festzuhalten, dass im Großen und Ganzen betrachtet zwischen Mikrozensuspanel und Beschäftigtenstichprobe keine gravierenden Unterschiede feststellbar sind, obwohl Berufswechsel im Mikrozensuspanel nur bei räumlich immobilen Ausbildungsabsolventen beobachtet werden können. In Anbetracht des trotz Kumulation der Absolventen von drei Übergängen geringen Stichprobenumfangs sind allerdings die Stichprobenfehler für weitergehende differenzierte Analysen zu groß. Dies betrifft einerseits die Unterscheidung von Gruppen mit unterschiedlichen Risiken des Berufswechsels, wenn nur die Daten des Mikrozensuspanels verwendet werden. Andererseits wird es aufgrund des Stichprobenfehlers auch schwierig, die Befunde auf Basis räumlich mobiler Absolventen des Mikrozensuspanels durch Vergleiche mit Ergebnissen der Beschäftigtenstichprobe genauer zu validieren.

## **6.5 Zusammenfassung**

Auszubildende bzw. Absolventen des dualen Systems sind überdurchschnittlich räumlich mobil. Zur Beantwortung der Frage, ob Analysen dieser Teilpopulation aufgrund der Ausfälle im Mikrozensuspanel verzerrt sind, wurden log-lineare Modelle verwendet. Zur Überprüfung der Ausfallannahmen und der Modellergebnisse wurde die Beschäftigtenstichprobe genutzt, deren Verlaufsangaben nicht durch Fort- und Zuzüge verzerrt sind. Hierfür war es notwendig, möglichst gut vergleichbare Daten zu konstruieren, da andernfalls Selektivitätseffekte durch andere Differenzen zwischen den Datenquellen überlagert werden.

Die deskriptiven Vergleiche ergaben, dass Auszubildende des dualen Systems im Mikrozensus im Vergleich zur Beschäftigtenstichprobe systematisch untererfasst sind, wobei die Abweichungen 1996/97 zwölf Prozent betragen. Die Gründe für die Untererfassung der Auszubildenden konnten zwar nicht aufgeklärt werden, doch kann bei den Analysen zum Ausfall weitestgehend ausgeschlossen werden, dass deren Ergebnisse durch Klassifikationsfehler konfundiert sind.

Die Analysen zum Problem selektiver Ausfälle konzentrierten sich auf Auszubildende des Jahres 1996 und den Übergang bis 1997, bei dem 19 Prozent der Auszubildenden infolge von Wegzügen aus dem Auswahlbezirk des Mikrozensus ausgefallen sind. Überdurchschnittliche Ausfälle sind bei Frauen und für Ausbildungsstätten in Ostdeutschland (inkl. Berlin) festzustellen. Der Ausfall hängt auch eng mit dem Alter und der Beschäftigungsdauer im Ausbildungsbetrieb zusammen.

Im Unterschied zu Auszubildenden und Absolventen zeigen sich deutlich stärkere Ausfallrisiken von Ausbildungsabbrechern mit einer längeren Beschäftigungsdauer im Ausbildungsbetrieb. Damit liegen nicht ignorierbare, mit dem Statuswechsel selbst verbundene Ausfälle vor. Das Muster von Verteilungsabweichungen zur Beschäftigtenstichprobe korreliert mit der Ausfallrate des Mikrozensuspanels und bildet die Tendenz des „Bleibens der Erfolgreichen“ bzw. des „Fortzugs der Ausbildungsabbrecher“ ab.

Weil die Ausfallrisiken für Auszubildende, erfolgreiche Ausbildungsabsolventen und Ausbildungsabbrecher nicht gleich sind, können mit Gewichtungsfaktoren, die auf der Annahme bedingt zufälliger Ausfälle (MAR) beruhen, keine wirksamen Korrekturen der Stichprobenselektivität erreicht werden (vgl. dazu auch die Simulationsergebnisse in Lehmann 2004). Entsprechende Tests zeigen deutlich, dass die Verteilungen der so gewichteten Daten räumlich mobiler Personen des Mikrozensuspanels von den Verteilungen der Beschäftigtenstichprobe abweichen und zu anderen inhaltlichen Ergebnissen führen können.

Zur Frage der Chancen einer Beschäftigung im erlernten Beruf nach Abschluss der Ausbildung wurden deskriptive Analysen der räumlich mobilen Absolventen des Mikrozensuspanels durchgeführt. Während nach dem Mikrozensuspanel rund 17 Prozent der erwerbstätigen Ausbildungsabsolventen einen Berufswechsel aufweisen, liegt der Anteil in der Beschäftigtenstichprobe bei rund 15 Prozent. Die Differenz kann mit der Selektion räumlich mobiler Absolventen, aber auch mit der in beiden Datenquellen unterschiedlichen Berufscodierung zusammenhängen. Berücksichtigt man den vergleichsweise hohen Stichprobenfehler des Mikrozensuspanels, können keine gravierenden Unterschiede zum Berufswechsel festgestellt werden.





## 7 Zusammenfassung und Ausblick

Für Deutschland wurde unter anderem im Gutachten der Kommission zur Verbesserung der informationellen Infrastruktur (KVI 2001: 15, 49) und im ersten Armuts- und Reichtumsbericht der Bundesregierung (BMA 2001: 37) im internationalen Vergleich ein Defizit an Längsschnittdaten festgestellt. Um diesen Nachholbedarf aufzuholen und bessere Möglichkeiten für dynamische Analysen sowie für Analysen des sozialen Wandels zu schaffen, hat die KVI u. a. empfohlen, den Mikrozensus weiterzuentwickeln und dabei insbesondere die rechtlich seit 1996 mögliche Zusammenführung von Querschnittsdaten des Mikrozensus zu einem Panel zu nutzen. Da die Haushalte eines Auswahlbezirks im Mikrozensus vier Jahre lang befragt werden und jedes Jahr ein Viertel der Auswahlbezirke (Klumpen) ausgetauscht wird, können rotierende Panelstichproben gebildet werden, die einen Zeitraum von zwei bis vier Jahren abdecken.

Im Vergleich zu anderen Panels sind dies zwar kurze Zeiträume, jedoch stehen dieser Einschränkung andere Vorteile des Mikrozensus gegenüber. Aufgrund des hohen Auswahlssatzes von einem Prozent sind, selbst wenn nur maximal ein Viertel der Auswahlseinheiten zur Verfügung steht, differenzierte Analysen durchführbar. Damit wird es möglich, ökonomische und soziale Strukturen und ihre Veränderungen in tiefer sachlicher Gliederung zu analysieren, auch für kleine Bevölkerungsgruppen verlässliche Aussagen zu machen und den sozialen und wirtschaftlichen Wandel differenzierter als bisher zu beobachten. Infolge der Teilnahmepflicht der ausgewählten Haushalte ist beim Mikrozensus der sonst in vielen Panels problematische Befragungsausfall bei der ersten Erhebung (Martin et al. 2006) nahezu vernachlässigbar.

Die Durchführung des Mikrozensus als Flächen- bzw. Klumpenstichprobe gewährleistet des Weiteren, dass in den Auswahlbezirken „automatisch“ alle Populationsveränderungen erfasst werden. Im Unterschied zu anderen Panels sind damit keine Aufstockungen durch Ergänzungstichproben nötig, die z. B. zur Erfassung der seit Beginn eines Panels hinzugekommenen Migranten durchgeführt werden. Nicht zuletzt fallen durch die Zusammenführung bereits bestehender Querschnittsdaten des Mikrozensus keine zusätzlichen Erhebungskosten an. Somit kann ein Mikrozensuspanel vergleichsweise kostengünstig zur Deckung des Bedarfs an längsschnittlichen Informationen beitragen.

Nach dem Prinzip der Flächenstichprobe werden die aus dem Auswahlbezirk wegziehenden Haushalte und Personen nicht weiter befragt, sondern durch die zuziehenden Haushalte und Personen ersetzt. Verlaufsanalysen des Mikrozensuspanels sind deshalb mit dem Problem konfrontiert, dass aufgrund von Ausfällen bei der Schätzung von Übergangswahrscheinlichkeiten nur die Daten von räumlich Immobilen zur Verfügung stehen. Wenn sich aber die Gruppen der räumlich mobilen und immobilen Personen in ihren Merkmalsverteilungen unterscheiden oder wenn das Risiko eines Ausfalls mit der Merkmalsveränderung zusammenhängt, ist die Stichprobe der räumlich Immobilen nicht mehr als einfache Zufallsauswahl der Gesamtheit zu betrachten. Das Ausmaß einer möglichen Verzerrung hängt vom Anteil der ausgefallenen Personen an der betrachteten Population und der Kovariation der interessierenden Merkmale mit dem Ausfall ab. Das Selektivitätsrisiko kann also je nach Fragestellung und Zielpopulation unterschiedlich hoch sein.

Die mit dem Ausfall verbundenen Probleme waren für ausgewählte zentrale Themenbereiche des Mikrozensus anhand der ersten Stichprobe des Mikrozensuspanels 1996-1999 zu evaluieren. Vor dem Hintergrund erheblicher Defizite längsschnittlicher Informationen in der empirischen Bildungsforschung wurden deshalb diese Selektivitätsfragen am Beispiel von drei Übergängen

untersucht: Dem Auszug aus dem Elternhaus, dem Besuch der gymnasialen Oberstufe bis zum Abitur und der beruflichen Ausbildungen des dualen Systems bis zum Abschluss.

Betrachtet man die Stichprobenselektivität bzw. potenzielle Verzerrung als allgemeines Problem fehlender Daten (Little und Rubin 2002), kommen verschiedene Verfahren zur Aufklärung von Ausfallursachen und zur Korrektur von Ausfällen in Frage. Für die im Mikrozensus dominierenden kategorialen Angaben wurden in dieser Arbeit spezielle log-lineare Modelle bzw. log-lineare Pfadanalysen verwendet. Diese zur Klasse der Selektionsmodelle gehörenden Verfahren sind in der empirischen Sozialforschung bisher wenig eingesetzt worden (Pötter 2006). Mit diesen Modellen wird die gemeinsame Wahrscheinlichkeit für die unter inhaltlichen Aspekten interessierenden Variablen und den Ausfall geschätzt. Durch die im Selektionsmodell verwendete Faktorisierung der gemeinsamen Wahrscheinlichkeit ergeben sich als Teilmodelle ein Strukturmodell der substanzwissenschaftlich relevanten Variablen und ein Ausfallmodell. Damit wird es möglich zu untersuchen, wie die Ausfälle mit den Analysevariablen zusammenhängen. Des Weiteren können damit modellbasierte Schätzungen für die Gesamtdaten, d. h. inklusive den Ausfällen, für die interessierenden Zusammenhänge vorgenommen werden.

Allerdings hängt die Gültigkeit der Ergebnisse davon ab, ob die bei der Modellierung der Zusammenhänge zwischen Ausfällen und interessierenden Variablen zugrunde gelegten Annahmen zutreffen. Ausschließlich auf der Grundlage des Mikrozensuspanels kann die Gültigkeit der Annahmen nicht begründet werden; letztlich kann man darüber nur spekulieren (Pötter 2006: 199). Zur Validierung sind a priori Kenntnisse über das Ausfallgeschehen bzw. externe Daten zur Population unerlässlich (Allison 2002). Hierfür wurden zwei verschiedene Referenzdaten verwendet: Für den Übergang vom Besuch der gymnasialen Oberstufe bis zum Abitur wurden Querschnittsergebnisse der amtlichen Schulstatistik als Populationsdaten genutzt. Die Referenzdaten für die Übergänge von Auszubildenden des dualen Systems stammen aus der IAB-Beschäftigtenstichprobe (IABS-R01), einer Stichprobe der Längsschnittpopulation sozialversicherungspflichtig Beschäftigter.

Insofern bei einem Panel auf die Angaben aus vorherigen Befragungen zurückgegriffen werden kann, liegen auch neue Informationen zur Antwortkonsistenz im Mikrozensus vor. Diese Informationen erlauben die partielle Untersuchung der Datenqualität und die Bearbeitung von Fragen der Verwendbarkeit des Mikrozensus für Längsschnittanalysen, die über die Problematik der Stichprobenselektivität und ihrer möglichen Korrektur hinausgehen.

Im Folgenden werden die wichtigsten Ergebnisse zusammengefasst. Sie beziehen sich auf die zentrale Frage selektiver Ausfälle und die Validitätsaspekte der Bildungsangaben des Mikrozensus im weiteren Sinne. Obwohl der Mikrozensus als Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten (partielle Rotation) durchgeführt wird, ist das Erhebungsdesign und Fragenprogramm des Mikrozensus gegenwärtig noch weitestgehend auf Querschnittsergebnisse ausgerichtet. Vor dem Hintergrund der dadurch restringierten Möglichkeiten für Verlaufsanalysen werden abschließend Fragen aufgeworfen, wie das Potenzial für Längsschnittanalysen künftig besser ausgeschöpft und vergrößert werden könnte.

## 7.1 Stichprobenselektivität bei Bildungsverlaufsanalysen

Die für die Analyse von Bildungsverläufen relevante Altersgruppe ist insgesamt räumlich hoch mobil und fällt damit überdurchschnittlich häufig im Mikrozensuspanel aus. Insbesondere im Alter von 18 bis 30 Jahren verlassen viele Jugendliche den weiter befragten Haushalt ihrer Eltern, zum Beispiel aus ausbildungs- oder berufsbedingten Gründen. Bis 1999 fallen knapp 30 Prozent

der 16- bis 19-jährigen Schüler der gymnasialen Oberstufe des Jahres 1996 aus. In Bezug auf die Umzugshäufigkeit unterscheiden sich zwar die Gymnasiasten nur wenig vom Durchschnitt aller Personen im Mikrozensuspanel. Allerdings korrelieren ihre Ausfall- bzw. Umzugsanteile tendenziell mit dem Alter, d. h. mit der zeitlichen Nähe zum interessierenden Abschluss. Stärker trifft dies für die 15- bis 24-Jährigen zu, die 1996 eine berufliche Ausbildung im dualen System absolviert haben. Von ihnen fallen bis 1999 mit rund 47 Prozent weit mehr aus als bei den Gymnasialschülern. Es war deshalb zu vermuten, dass Analysen junger Erwachsener auf Basis des Mikrozensuspanels infolge des Ausfalls räumlich mobiler Personen zu verzerrten Ergebnissen führen.

#### *Auszug aus dem Elternhaus*

Die Ausgangshypothese, dass die hohe räumliche Mobilität der Jugendlichen zu verzerrten Ergebnissen bei Analysen dieser Teilpopulation führt, konnte nach einer Diskussion theoretischer Ansätze zum Auszug aus dem Elternhaus als Spezialfall räumlicher Migration und der Ergebnisse von Verlaufsanalysen zum Auszugsverhalten weiter konkretisiert werden (siehe Kap. 4). Im Hinblick auf die objektive Gelegenheitsstruktur als eine Rahmenbedingung von Auszugsentscheidungen ist beispielsweise in den neuen Bundesländern aufgrund der dort angespannten Situation auf dem Arbeitsmarkt mit einer höheren Auszugsneigung als in Westdeutschland zu rechnen. In Bezug auf die sozioökonomische Lage der Herkunftsfamilie ist zu erwarten, dass Familien mit höherem Einkommen ihren Kindern ein komfortableres Wohnumfeld bieten können, das die Auszugsneigung verringert bzw. Auszüge später stattfinden lässt als bei Jugendlichen einkommensschwächerer Familien. Insbesondere für die Übergänge im Zusammenhang zwischen Ausbildung und Berufseinstieg hat die Diskussion des Forschungsstandes gezeigt, dass Auszüge aus dem Elternhaus nach einem Abschluss der Ausbildungsphase verstärkt stattfinden und dieser Anstieg mit der Verfügbarkeit über ein eigenes ausreichendes Erwerbseinkommen verbunden ist.

Für die empirische Analyse wurden hypothesengeleitet Faktoren gewählt, die den Auszug fördern bzw. hinausschieben, wie z. B. eigenes Einkommen vs. Abhängigkeit von finanziellen Transferleistungen der Familie. In den Verlaufsanalysen wurde der Verbleib 15- bis 26-Jähriger im Elternhaus untersucht. Die Stichprobe enthält Jugendliche, die 1996 als lediges Kind bei ihren Eltern wohnten ohne Personen, die mit dem gesamten Haushalt weggezogen sind. Die Ergebnisse entsprechen weitgehend denen anderer Studien. Sie deuten an, dass Längsschnittauswertungen des Mikrozensus noch weitere Analysemöglichkeiten zum Thema des Verbleibs Jugendlicher im Elternhaus und der Veränderung von Familienbeziehungen bieten, wie beispielsweise zu Übergängen der Lebensformen Ehe und nichteheliche Lebensgemeinschaft.

Bemerkenswert ist, dass die in Ostdeutschland im Vergleich zum früheren Bundesgebiet höheren Auszugswahrscheinlichkeiten hauptsächlich mit geschlechtsspezifischen Unterschieden und einem stärkeren Zusammenhang zwischen Auszug und Geschwisterzahl zusammenhängen. Nicht bestätigt werden konnte, dass Jugendliche früher von zu Hause ausziehen, wenn sie über ein eigenes Erwerbseinkommen verfügen. Dies spricht aber weniger gegen die Hypothese, sondern dürfte vielmehr mit dem engen statistischen Zusammenhang der Quelle des überwiegenden Lebensunterhalts als Einkommensindikator mit Variablen der Bildungs- und Erwerbsbeteiligung begründet sein.

In Bezug auf die Ausfallproblematik sind zwei Teilergebnisse der Verlaufsanalysen von besonderer Bedeutung. Der Befund, dass sich die geschätzten Auszugswahrscheinlichkeiten für Besucher allgemein bildender Schulen und für Hochschulbesucher in Westdeutschland statistisch nicht signifikant unterscheiden, deutete darauf hin, dass Analysen der Übergänge der Besucher der gymnasialen Oberstufe auf Basis räumlich immobiler Personen nicht wesentlich verzerrt sein dürften. Dagegen waren bei betrieblichen Ausbildungen Zusammenhänge zwischen Ausfall und

Statuswechsel zu vermuten, da Ausbildungsepisoden nach der allgemein bildenden Schule, Beendigungen des Berufsschulbesuchs sowie Betriebswechsel zum statistisch signifikanten Anstieg der Auszugswahrscheinlichkeiten führen.

Die Frage, ob Wegzüge bzw. Ausfälle mit einem Wechsel des Ausbildungsstatus einhergehen, lässt sich mit den Verlaufsanalysen nicht beantworten. Hierfür wäre es nötig, Zielzustände nach dem Umzug zu betrachten. Dazu liegen aber im Mikrozensus keine Daten vor. Die Analysen zum Auszug junger Erwachsener aus dem Elternhaus gaben dennoch wichtige Hinweise zur Selektivitätsproblematik, die in die Mobilitäts- bzw. Ausfallmodelle bei den anschließenden Analysen eingeflossen sind.

#### *Vom Besuch der gymnasialen Oberstufe bis zum Abitur*

Bei der Untersuchung des Verlaufs vom Besuch der gymnasialen Oberstufe bis zum Abschluss mit dem Abitur konnte mit dem Mikrozensuspanel insofern Neuland betreten werden, als bisher mit den Querschnittsdaten des Mikrozensus entweder nur Analysen zum Schulbesuch bzw. zur Bildungsbeteiligung oder zum Bildungsabschluss möglich waren. Die soziale Selektivität des erfolgreichen Abschlusses der gymnasialen Oberstufe ist auf Individualebene mit den Daten der amtlichen Schulstatistik nicht untersuchbar. Auch sonst liegen kaum Ergebnisse dazu vor. Der Evaluation des Mikrozensuspanels kam daher eine besondere Bedeutung zu. Zu klären, ob selektive Ausfälle im Mikrozensuspanel vorliegen und in welcher Weise diese gegebenenfalls korrigiert werden können, war auch deshalb interessant, weil dann diese Forschungsfrage mit den Paneldaten untersuchbar wäre.

Bis einschließlich 2004 wurden im Mikrozensus die Angaben zum Schulbesuch und Bildungsabschluss jährlich im Frühjahr erfragt. Die Berichtswoche lag i. d. R. in der letzten Aprilwoche. Im April steht aber ein erfolgreicher Abschluss der gymnasialen Oberstufe noch nicht fest, sondern erst zum Ende des Schuljahres. Deshalb kann mit den zu einem Panel zusammengeführten Querschnittsdaten erst in der Befragung ein Jahr darauf ermittelt werden, welche Schüler das Abitur erreicht haben, sofern sie bis dahin nicht umgezogen sind. Die Analysestichprobe der Schüler, die zwischen 1996 und 1999 die gymnasiale Oberstufe besucht haben, umfasst pro Übergang 1996/97, 1997/98 und 1998/99 jeweils rund 1.700 Schüler. Konzentriert man sich auf die Risikopopulation der 16- bis 19-Jährigen des Jahres 1996 und ihre Chancen des erfolgreichen Abschlusses bis 1999, stehen rund 1.300 Schüler für die Analysen zur Verfügung. Für rund 18 Prozent der Risikopopulation kann aufgrund von Fortzügen nicht ermittelt werden, ob sie das Abitur erreicht haben.

Unabhängig von der Ausfallproblematik würde man im Allgemeinen annehmen, dass die bis zur elften Klasse der gymnasialen Oberstufe aufgerückten Schüler überwiegend bis zum Abitur dort verbleiben. Unter theoretischen Gesichtspunkten und in Anlehnung an das auf Boudon (1974) zurückgehende handlungstheoretische Modell zur Bildungsungleichheit ist die soziale Selektion in den frühen Phasen der Bildungslaufbahn am ausgeprägtesten, d. h. beim Übergang von der Primarstufe zur Sekundarstufe I. Für zeitlich nachgelagerte Stufen wie die der gymnasialen Oberstufe ist von einer geringeren sozialen Ungleichheit auszugehen.

Zur zentralen Frage, in welchem Maße bei der Ermittlung eines Abschlusses selektive Ausfälle vorliegen, war nach den Analysen zum Auszug aus dem Elternhaus mit geringen bzw. korrigierbaren Verzerrungen zu rechnen. Die ersten deskriptiven Analysen zeigten dagegen einerseits eine unerwartet hohe Fluktuation in den Antworten zum Schulbesuch. Insbesondere finden sehr häufig Wechsel in berufliche Schulen statt. Andererseits ergaben sich unplausibel niedrige Quoten des Abschlusses mit dem Abitur. Selbst bei den räumlich immobilen Schülern, die 1996 im Alter von 17 Jahren überwiegend die elfte Klasse besucht haben dürften, betragen die Abschluss-

quoten nur rund 70 Prozent. Die daraufhin vorgenommenen ausführlichen Analysen zur Datenqualität der Bildungsangaben im Mikrozensus (siehe Abschnitt 5.3) zeigten im Vergleich zur amtlichen Schulstatistik eine erhebliche Übererfassung der gymnasialen Oberstufenschüler. Schüler der Sekundarstufe I und Schüler beruflicher Schulen sind im Mikrozensus teilweise als Besucher der gymnasialen Oberstufe fehlklassifiziert. Dies vergrößert bei der Berechnung der Abschlussquote den Umfang der Risikopopulation im Nenner und führt zu einer gravierenden Unterschätzung der Absolventenquote. In den log-linearen Modellen zur Stichprobenselektivität konnten die Klassifikationsfehler allerdings nicht berücksichtigt werden, da die vorliegenden Daten keine Identifikationsmöglichkeiten bieten. Dafür wären Wiederholungsbefragungen hilfreich (siehe hierzu Abschnitt 5.5)

Hinsichtlich der Panelausfälle hängt einerseits die berufliche Stellung des Familienvorstands der Schüler sehr eng mit dem Ausfall zusammen. Insbesondere Arbeiterkinder fallen überdurchschnittlich häufig aus. Dies spricht für die Annahme eines bedingt zufälligen Ausfalls („missing at random“, MAR), der von der beruflichen Stellung des Familienvorstands der Schüler abhängt. Damit kann zugleich die überdurchschnittliche räumliche Mobilität von Arbeiterhaushalten korrigiert werden, die in den Verlaufsanalysen zum Auszug aus dem Elternhaus nicht erkennbar war, da die räumliche Mobilität ganzer Haushalte unberücksichtigt blieb. Andererseits ergaben sich aber statistisch signifikante Zusammenhänge des Ausfalls mit dem (nicht beobachteten) Schulbesuch bzw. Abschluss. Überlegungen zur Plausibilität dieser gegensätzlichen Ausfallannahmen sprachen zwar für die Annahme bedingt zufälliger Ausfälle. Alleine auf Basis der Daten des Mikrozensuspanels war jedoch keine Entscheidung möglich.

Auch wenn in der amtlichen Bildungsstatistik keine Verlaufsangaben und keine Angaben über die berufliche Stellung der Eltern vorliegen, ist es dennoch näherungsweise möglich, die Plausibilität der Modelle durch den Vergleich von Querschnittsergebnissen des Mikrozensuspanels mit den Absolventenzahlen der Bildungsstatistik zu prüfen. Gilt die Annahme bedingt zufälligen Ausfalls, sollte insgesamt betrachtet eine Ausfallkorrektur durch die Gewichtung mit dem Kehrwert der Wahrscheinlichkeit für die räumliche Immobilität bzw. Antwortwahrscheinlichkeit erreicht werden.

Zur Validierung der Modellbefunde wurde die mit dem Mikrozensuspanel geschätzte Absolventenzahl allgemein bildender Schulen mit Allgemeiner Hochschulreife des Jahres 1996 mit den Ergebnissen der Bildungsstatistik verglichen. Verwendet man bei der Analyse der räumlich immobilen Schüler für die Gewichtung die nach der beruflichen Stellung des Familienvorstandes sowie nach West- und Ostdeutschland differenzierten Antwortwahrscheinlichkeiten, kann eine gute Anpassung der mit dem Mikrozensuspanel geschätzten Zahl der Abiturienten an die Ergebnisse der Bildungsstatistik erreicht werden.

Zusammenfassend erscheinen somit die durch Panelausfälle bei Verlaufsanalysen entstehenden Selektivitätsprobleme mithilfe von Gewichtungsfaktoren korrigierbar. Dagegen ließe sich einwenden, dass diese aus dem Vergleich des Mikrozensuspanels mit Populationsergebnissen resultierende Einschätzung im Grunde auf Annahmen zur Vergleichbarkeit beruht, die aufgrund von Klassifikationsfehlern nicht haltbar sind. Beachtet man aber, dass sich mit einer Ausfallgewichtung, die den Annahmen nicht ignorierbarer Ausfälle entspricht, größere Abweichungen zur Bildungsstatistik ergeben würden, gibt es gute Gründe dafür, die Annahme bedingt zufälliger Ausfälle zu akzeptieren.

Vorausgesetzt es liegen keine weiteren systematischen Fehler vor, wäre es auf der Grundlage dieser Ergebnisse möglich, die oben angesprochene Frage der sozialen Selektivität des erfolgreichen Abschlusses der gymnasialen Oberstufe zu beantworten. Die mit entsprechender Ausfallkorrektur geschätzte Abschlussquote ist aber wegen der Fehlklassifikationen unplausibel niedrig



(siehe Abb. 5.11). Somit ist die Frage zur sozialen Selektivität des Bildungserfolgs in der gymnasialen Oberstufe auf Basis des Mikrozensuspanels gegenwärtig nicht beantwortbar. Die Ursache dafür liegt jedoch nicht in der Stichprobenselektivität – diese ist unter den oben genannten Vorbehalten korrigierbar –, sondern bei den hier aufgedeckten Klassifikationsfehlern der Angaben zum Schulbesuch im Mikrozensus.

### *Berufliche Ausbildungsverläufe bis zum Übergang ins Erwerbsleben*

Ergänzend zu den Bildungsverläufen in der gymnasialen Oberstufe sollte mit Analysen zur beruflichen Ausbildung geklärt werden, in welcher Weise mit dem Mikrozensuspanel der in diesem Themengebiet dringende Bedarf an längsschnittlichen Informationen abgedeckt werden kann. Dies hängt wesentlich vom Ausmaß selektiver Ausfälle ab. Im Gegensatz zur gymnasialen Oberstufe waren, wie die Analysen zum Auszug aus dem Elternhaus gezeigt haben, aufgrund eines Anstiegs der Auszugswahrscheinlichkeit nach dem Ende von Ausbildungsepisoden sowie im Zusammenhang mit Betriebswechseln selektive Ausfälle zu erwarten.

Im Unterschied zu den Übergängen der Oberstufenschüler lagen durch die Verwendbarkeit der Beschäftigtenstichprobe für die Selektivitätsprüfungen günstigere Bedingungen vor. Diese Daten enthalten als eine Stichprobe der Längsschnittpopulation der sozialversicherungspflichtig Beschäftigten alle Auszubildenden des dualen Systems. Obgleich die Beschäftigtenstichprobe keine Informationen zum Umzug und Wohnort enthält, sind die Ausbildungs- und Erwerbsverläufe nicht durch Fort- und Zuzüge verzerrt.

Durch eine für beide Datenbasen vergleichbare Abgrenzung der Auszubildenden des dualen Systems wurde es möglich, eventuell bei einem Vergleich störende Differenzen zwischen den unterschiedlich generierten Daten weitestgehend auszuschließen. Somit lässt sich indirekt aus den Verteilungsabweichungen auf die im Mikrozensuspanel nicht beobachtbaren Effekte räumlicher Mobilität schließen. Zu beachten ist, dass die Referenzdaten der Beschäftigtenstichprobe bei den nicht meldepflichtigen Merkmalen (z. B. Bildungs- und Berufsangaben) Codier- und Klassifikationsfehler enthalten können. Da aber bei allen meldepflichtigen Merkmalen, speziell den Beschäftigungszeiten von Auszubildenden, von einer sehr hohen Genauigkeit ausgegangen werden kann, bietet die Verwendung der Beschäftigtenstichprobe entscheidende Vorteile. Der Einsatz dieser Stichprobe der Längsschnittpopulation erlaubt es insbesondere, bei den Selektionsmodellen die Annahme gleicher Zusammenhänge zwischen den Analysevariablen für das Mikrozensuspanel und die Referenzdaten zu modellieren. Auch wenn in der Beschäftigtenstichprobe nur wenige Merkmale vorliegen, die für eine Erklärung räumlicher Mobilität relevant sind, verbessert dies die Identifikation von Ausfallzusammenhängen in ganz entscheidender Weise.

Die Analysen konzentrierten sich auf Auszubildende des Jahres 1996 und den Übergang bis 1997, bei dem 19 Prozent der Auszubildenden infolge von Wegzügen aus dem Auswahlbezirk des Mikrozensus ausgefallen sind. Erste deskriptive Vergleiche von Bestands- bzw. Querschnittergebnissen ergaben, dass Auszubildende des dualen Systems im Mikrozensuspanel im Vergleich zur Beschäftigtenstichprobe systematisch untererfasst sind. Die Abweichungen betragen 1996 und 1997 immerhin zwölf bzw. 14 Prozent. Ob wie beim Besuch der gymnasialen Oberstufe Klassifikationsfehler vorliegen, konnte zwar nicht geklärt werden. Doch konnte geprüft werden, dass diese Untererfassung die Auszugsanalysen nicht beeinträchtigt.

In den Selektionsmodellen wurden Übergänge der Auszubildenden des dualen Systems in Erstausbildung im Jahre 1996 untersucht, und zwar ob sie 1997 noch in Ausbildung sind, einen Ausbildungsabschluss erreicht oder die Ausbildung abgebrochen haben. Wie zu erwarten, hängt der erreichte Status maßgeblich von der Ausbildungsdauer ab. Andere erklärende Variablen weisen schwächere Zusammenhänge auf. So brechen ausländische Jugendliche im Vergleich zu

Deutschen häufiger die Ausbildung ab. Auszubildende mit Abitur haben größere Chancen die Ausbildung abzuschließen als Auszubildende ohne Abitur.

In Bezug auf Ausfallmuster zeigten sich keine statistisch signifikanten Zusammenhänge für den allgemeinen Bildungsabschluss sowie die Staatsangehörigkeit. D. h. Auszubildende mit Abitur und Ausländer unterscheiden sich in dieser Hinsicht nicht von Personen ohne Abitur bzw. Deutschen. Erwartungsgemäß wurden höhere Ausfallwahrscheinlichkeiten für ältere Auszubildende, Frauen, Auszubildende ostdeutscher Ausbildungsstätten (inkl. Berlin) und bei längerer Beschäftigungsdauer im Ausbildungsbetrieb ermittelt. Von besonderer Bedeutung sind die Ergebnisse zum Alter und zur Beschäftigungsdauer. Sie zeigen, dass Personen im Mikrozensuspanel infolge von Umzügen mit zunehmender zeitlicher Nähe zum Ausbildungsabschluss ausfallen.

Darüber hinaus ergaben sich im Vergleich zu Auszubildenden und Ausbildungsabsolventen deutlich stärkere Ausfallrisiken von Ausbildungsabbrechern in Kombination mit einer längeren Beschäftigungsdauer im Ausbildungsbetrieb. Es kann also nicht davon ausgegangen werden, dass die Ausfallrisiken für Auszubildende, erfolgreiche Ausbildungsabsolventen und Ausbildungsabbrecher gleich sind, die in der Annahme von bedingt zufälligen Ausfällen zugrunde gelegt werden. Wie bereits aufgrund der Verlaufsanalysen zum Auszug aus dem Elternhaus zu erwarten war, liegen damit nicht ignorierbare Ausfälle vor.

Während aber verschiedene Überlegungen zum Auszug aus dem Elternhaus einen Zusammenhang zwischen dem erfolgreichen Ausbildungsabschluss und dem Umzug bzw. Ausfall nahelegen, zeigt sich hier hinsichtlich des im Vergleich zu den Ausbildungsabbrechern geringeren Ausfallrisikos von Ausbildungsabsolventen die Tendenz des „Bleibens der Erfolgreichen“ bzw. des „Fortzugs der Ausbildungsabbrecher“. Dieses Ergebnis steht vermutlich mit der Mitte bis Ende der 1990er Jahre schwierigen Situation auf dem Arbeitsmarkt in Verbindung. Vor dem Hintergrund rückläufiger Chancen der Übernahme der Auszubildenden in ein reguläres Beschäftigungsverhältnis ist anzunehmen, dass es für erfolgreiche Absolventen vorteilhaft ist, zumindest kurz- bis mittelfristig im Ausbildungsbetrieb zu bleiben. Zudem ist davon auszugehen, dass Ausbildungsabsolventen eher ein Übernahmeangebot erhalten als Ausbildungsabbrecher. Wenn bei Betriebswechseln die Wahrscheinlichkeit von Umzügen steigt, ist bei Ausbildungsabbrechern mit einem höheren Ausfallrisiko als bei erfolgreichen Absolventen zu rechnen.

Bei den deskriptiven Analysen der Chancen der Beschäftigung im erlernten Beruf wurden im Mikrozensuspanel nur räumlich immobile Ausbildungsabsolventen ohne Ausfallkorrektur betrachtet. Es waren keine gravierenden Unterschiede zur Beschäftigtenstichprobe feststellbar. Aufgrund der Ergebnisse eher geringer Ausfallrisiken von Ausbildungsabsolventen ist dies wenig überraschend.

### *Korrekturmöglichkeiten bei nicht ignorierbaren Ausfällen*

Der obige Befund nicht ignorierbarer Ausfälle (NINR) ist mit der Konsequenz verbunden, dass bei Analysen räumlich immobiler Personen zum Verlauf und Abschluss einer beruflichen Ausbildung im dualen System die dem NINR-Modell entsprechenden Antwortwahrscheinlichkeiten bei der Ausfallkorrektur bzw. Gewichtung berücksichtigt werden müssen. Allerdings ist mit diesem Vorgehen ein Mehraufwand verbunden. Man benötigt nicht nur Hypothesen zu den substanzwissenschaftlichen Fragestellungen, sondern auch zur räumlichen Mobilität bzw. zum Ausfall. Zudem gehören die statistischen Verfahren nicht zum Standardrepertoire der empirischen Sozial- und Wirtschaftsforschung (vgl. Pötter 2006).

Aus pragmatischer Sicht stellt sich deshalb die Frage, in welchem Maße mit Verzerrungen zu rechnen ist, falls die bereits im Scientific Use File des Mikrozensus bereitgestellten Gewichtungsvariablen verwendet werden (siehe Unterabschnitt 2.1.2). Um mit diesen Gewichten die Verzer-



rungen korrigieren zu können, die ggf. durch räumliche Mobilität entstehen, sollte kein Zusammenhang zwischen dem interessierenden Merkmal und dem Ausfall bestehen.

Da im Fall der beruflichen Ausbildungsverläufe die Annahme bedingt zufälligen Ausfalls (MAR) verworfen werden muss, können im Grunde genommen mit den darauf bzw. den auf der Annahme homogener Antwortwahrscheinlichkeiten (RHG) basierenden Gewichtungsfaktoren keine wirksamen Ausfallkorrekturen erreicht werden. Teilweise wird aber angenommen, dass Ausfälle dennoch mit solchen Gewichten korrigiert werden können, auch wenn die Ausfallannahmen verletzt sind (Little und Rubin 2002: 346; Schafer und Graham 2002: 164).

In dieser Hinsicht deuten bisherige Erfahrungen darauf hin, dass bei Fragen von Übergängen des Erwerbsstatus bereits eine Gewichtung mit dem Kehrwert der Wahrscheinlichkeit für räumliche Immobilität nach Geschlecht und Alter deutliche Korrekturen der Verzerrungen bewirken kann (Rendtel 2005).<sup>76</sup> Wirksame Korrekturen der Ausfälle im Mikrozensuspanel durch Gewichtungen konnten auch für Tätigkeits- und Betriebswechsel sowie Statusübergänge im Bereich Partnerschaft und Familie festgestellt werden (Konold 2005).

Jedoch trifft dies für die hier untersuchten Statusübergänge von Auszubildenden des Jahres 1996 nicht zu. Die deskriptiven Analysen und statistischen Modelle zeigten, dass die Verteilungen gewichteter Daten räumlich immobiler Personen des Mikrozensuspanels von den Verteilungen der Beschäftigtenstichprobe abweichen und sogar zu anderen inhaltlichen Ergebnissen führen können. Diese Ergebnisse sind allerdings spezifisch für das Mikrozensuspanel und diese Fragestellung. Für allgemeine Schlussfolgerungen wären systematische Simulationen notwendig (vgl. Lehmann 2004; Rässler 2000).

Die Gründe für das Scheitern der vorliegenden Gewichtungsvariablen bei der Korrektur von Verzerrungen sind einerseits in der Verletzung der MAR- bzw. RHG-Annahmen zu sehen, da die Ausfallrisiken für Auszubildende, erfolgreiche Ausbildungsabsolventen und Ausbildungsabbrecher gerade nicht gleich sind. Andererseits wurden bei der Konstruktion der Hochrechnungsfaktoren keine Übergänge des Erwerbs- bzw. Ausbildungsstatus, sondern nur demografische und familienstatistische Bestands- und Veränderungsdaten berücksichtigt. Bei Verwendung der im Mikrozensuspanel bereitgestellten Gewichtungsvariablen ist es also ratsam, darüber nachzudenken, ob ein Ausgleich der bei dem jeweils interessierenden Zusammenhang auftretenden Ausfälle durch die Gewichtungsvariablen möglich erscheint.

Als ein weiteres Ergebnis der Arbeit kann festgehalten werden, dass die für die spezielle Fragestellung ermittelten modellbasierten Gewichtungsfaktoren im Vergleich zur Verwendung der im File bereitgestellten Gewichtungsvariablen eine bessere Anpassung an Verteilungen der Referenzdaten ergaben. Dies trifft auch für die Analysen zu den Übergängen im Bereich der gymnasialen Oberstufe zu. Lassen sich (wenigstens teilweise) externe Referenzdaten oder Querschnittsdaten des Mikrozensuspanels für Überprüfungen der Stichprobenselektivität verwenden, kann bei einem gewissen Mehraufwand eine wirksame Reduktion der Verzerrungen erzielt werden.

## 7.2 Ausblick

Die Zusammenführung der Querschnittsangaben des Mikrozensus zu einem Panel bietet trotz der genannten Schwierigkeiten eine ganze Reihe von Vorteilen. Wie einleitend in diesem Kapitel erwähnt, erlaubt der Merkmalskanon dieser amtlichen Mehrthemenumfrage in Verbindung mit dem großen Stichprobenumfang und den fast vernachlässigbaren Ausfällen differenzierte Analy-

76 Diese Untersuchung basiert auf Daten des Sozioökonomischen Panels. Dabei wurden die gewichteten Daten räumlich immobiler Personen des Sozio-ökonomischen Panels mit den Gesamtdaten verglichen.

sen zur Struktur und zum Wandel sozialer und wirtschaftlicher Sachverhalte. Seit 2005 stehen zudem unterjährige Ergebnisse durch die Umstellung auf die kontinuierliche Befragung bereit. Mit den ab dem Erhebungsjahr 2005 erwarteten erhebungstechnischen Verbesserungen lassen sich voraussichtlich auf der Basis von Wohnungssubstichproben Querschnittsdaten mit zeitlich konsistenten Ordnungsnummern und Identifikatoren zu den Rotationsvierteln erstellen. Mithilfe dieser Identifikatoren können dann Forscher je nach Fragestellung und benötigtem Stichprobenumfang eigenständig Panels unterschiedlicher Beobachtungszeiträume konstruieren (Schimpl-Neimanns 2006a). Dieses Standardkonzept bietet in gleicher Weise wie bei den Mikrozensus-Querschnittsfiles flexible Auswertungsmöglichkeiten des Panels.

Mit der Erstellung des ersten Panelfiles des Mikrozensus können nun ähnlich zu den Rotationspanels des US-amerikanischen Current Population Survey (CPS) oder dem Labour Force Survey (LFS) des Vereinigten Königreiches Längsschnittanalysen durchgeführt werden. Wie bei den anderen Rotationspanels bleiben jedoch die durch die fehlende Weiterbefragung räumlich mobiler Personen entstehenden Ausfälle ein Problem.

Um das grundlegende Potenzial des Mikrozensus für die Längsschnittnutzung zu verbessern und die durch das gegenwärtige Stichprobendesign entstehenden Schwierigkeiten zu reduzieren oder ganz zu vermeiden, ist zu fragen, durch welche Maßnahmen dies erreicht werden kann. In dieser Hinsicht hat Müller (2003: 63) umfassende Revisionen vorgeschlagen:

*„Um den Mikrozensus in eine methodisch einwandfreie Panelerhebung von Haushalten und Personen umzugestalten, die dann den Wert des Mikrozensus für die Statistik wie für die Wissenschaft außerordentlich steigern würde, müsste das Stichprobenkonzept des Mikrozensus dringend entsprechend geändert werden. Die gegenwärtige Novellierung des Mikrozensusgesetzes wäre eine günstige Gelegenheit für einen solchen Quantensprung der amtlichen Statistik in die moderne Welt der Längsschnittdatenstatistik, mit der noch viele andere Vorteile verbunden wären. Die methodische Erneuerung des Mikrozensus für eine bessere Längsschnittnutzung müsste zu einem zentralen Entwicklungsziel des Mikrozensus werden.“*

Diese Möglichkeiten wurden jedoch bei der Verabschiedung des Mikrozensusgesetzes 2005 nicht genutzt. Das Stichprobenkonzept basiert nach wie vor auf der Flächenstichprobe und sieht keine Weiterbefragung wegziehender Personen und Haushalte vor. Da die rechtlichen Vorgaben des Mikrozensusgesetzes 2005 bis zum Jahre 2012 gelten, ist zu überlegen, wie das Potenzial für Längsschnittanalysen unter den gegebenen Restriktionen ausgeschöpft werden kann. Darüber hinaus ist zu diskutieren, unter welchen Bedingungen nach 2012 Längsschnittanalysen verbessert werden können.

Bis 2012 können die Selektivitätsprobleme ansatzweise auch ohne grundlegende Revision des Stichprobendesigns in zweifacher Weise reduziert werden. Zum einen wäre es hilfreich, bei zugezogenen Haushalten und Personen Angaben zum Status vor dem Zuzug zu erfragen. Zum anderen würde die Weiterbefragung einer Substichprobe der Wegziehenden Angaben zum Status nach dem Umzug liefern.

Weil diese Substichproben bis 2012 aus rechtlichen Gründen ohne Auskunftspflicht durchgeführt werden müssen, ist mit Einschränkungen der Aussagekraft aufgrund von Antwortausfällen zu rechnen. Bei der Retrospektivbefragung der in einen Auswahlbezirk hinzugezogenen Befragten ist es vorteilhaft, dass mit der Flächenstichprobe eine Querschnittsrepräsentativität verbunden ist. Allerdings können bei der Retrospektivbefragung Todesfälle und Umzüge ins Ausland nicht erfasst werden. Hinzu kommt, dass die Rückerinnerung an den ein Jahr zuvor eingenommenen Status verzerrt sein kann. Unter methodischen Aspekten besitzt die prospektive Weiterverfolgung

der aus einem Auswahlbezirk fortgezogenen Personen den Vorteil, dass keine Konditionierung auf „Überlebende“ stattfindet, sofern Wegzüge ins Ausland und Todesfälle identifizierbar sind. Dagegen dürfte die Ermittlung der neuen Adressen nicht immer einfach sein.

Trotz dieser Einschränkungen stünden aus beiden Verfahren Angaben der räumlich mobilen Personen zur Verfügung, die in verschiedener Weise genutzt werden können (vgl. Laaksonen und Chambers 2006; Rendtel 2005: 22). Erstens können mithilfe der Substichprobe Antwortwahrscheinlichkeiten für Statusübergänge räumlich mobiler Personen berechnet und für die Konstruktion von Korrekturgewichten eingesetzt werden. Dies erlaubt auch die Prüfung der Annahme bedingt zufälliger Ausfälle bzw. der Annahme homogener Ausfallgruppen, die den gegenwärtigen Korrekturgewichten für Längsschnittauswertungen zugrunde liegt. Zweitens kann die Substichprobe für die Modellierung von Statusübergängen räumlich mobiler Personen genutzt werden, um statistische Modelle für Statusübergänge zu schätzen. Vorausgesetzt, die Befragungsausfälle in der Substichprobe sind ignorierbar und der Stichprobenumfang ist ausreichend groß, kann sogar unter Umständen auf andere Referenzdaten zur Kontrolle von Ausfällen aufgrund räumlicher Mobilität verzichtet werden.

Allison erinnert aber daran, dass statistische Modelle und nachträgliche Korrekturverfahren als Hilfsmittel immer mit Einbußen verbunden sind:

*„The only really good solution to the missing data problem is not to have any. So in the design and execution of research projects, it is essential to put great effort into minimizing the occurrence of missing data.”* (Allison 2002: 2-3)

Diese generelle Lösung lässt sich nur durch die (partielle) Weiterbefragung wegziehender Haushalte und Personen lösen und setzt eine Modifikation des Stichprobendesigns nach 2012 voraus. Für eine optimale Längsschnittnutzung wäre eine Abkehr vom strikten Prinzip der Flächen- und Haushaltsstichprobe nötig. Insbesondere sind Regeln zur Weiterverfolgung der aus einem Haushalt fortziehenden Personen zu entwickeln. Dass die Weiterverfolgung über Landesgrenzen hinweg, also auch in der dezentral organisierten deutschen Statistik möglich scheint, zeigt das Beispiel des US-amerikanischen Survey of Income and Program Participation (SIPP 2001). Das SIPP ist eine rotierende Panelstichprobe, die ergänzend zum CPS durchgeführt wird. Im SIPP werden Personen über einen Zeitraum von vier Jahren im Abstand von vier Monaten befragt.

Beachtet man, dass der Mikrozensus in das System des EU Labour Force Survey eingebettet ist, die als Haushaltsstichproben mit dem primären Ziel von Querschnittauswertungen durchgeführt werden, setzen entsprechende Änderungen eine Abstimmung auf europäischer Ebene voraus. Zumindest könnte die testweise Durchführung der oben skizzierten Mobilitätssubstichproben zur Diskussion über das Stichprobenkonzept ab 2012 beitragen. Dieser Test liefert u. a. hilfreiche Informationen zum Aufwand der prospektiven Weiterverfolgung, der Güte von retrospektiven Befragungen der Zugezogenen sowie zum erforderlichen Stichprobenumfang.

Wenn im Rahmen des für 2011 geplanten registergestützten Zensus die bisherigen Qualitätsmängel der Einwohnermelderegister (siehe Statistische Ämter des Bundes und der Länder 2004) behoben werden können, käme ähnlich wie bereits beim österreichischen Mikrozensus (Haslinger und Kytir 2006) das Melderegister als Auswahlrahmen für Wohnungs-, Haushalts- und Personenstichproben in Frage. Die Ermittlung der neuen Adressen der aus dem Bestand fortgezogenen Personen dürfte mithilfe der Angaben des Einwohnermelderegisters leicht möglich sein.

Durch Änderungen des Erhebungsprogramms können ab 2012 weitere Verbesserungen von Längsschnittanalysen erreicht werden (vgl. Rendtel 2005: 25). Beispielsweise ist an Zusatzprogramme mit Fragen zu biografischen Angaben zu denken, die im österreichischen Mikrozensus bis 2003 (Berufs-, Bildungs-, und Geburtenbiografien; siehe Statistik Austria 2004) oder im CPS

(Einkommensverlauf und Heiratsbiografie; siehe Unicon 2007) durchgeführt wurden bzw. werden. Diese Angaben sowie weitere zeitbezogene Merkmale bieten vielfältige Möglichkeiten für Verlaufsanalysen und damit bedeutende Verbesserungen der Längsschnittnutzung des Mikrozensus.



# Literaturverzeichnis

- Afentakis, Anja, und Wolf Bihler, 2005: Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik* (10): 1039-1048.
- Allison, Paul D., 1982: Discrete-Time Methods for the Analysis of Event History. S. 61-98 in: Samuel Leinhardt (Hg.): *Sociological Methodology* 1982. San Francisco: Jossey Bass.
- Avenarius, Hermann, Hartmut Ditton, Hans Döbert, Klaus Klemm, Eckhard Klieme, Matthias Rürup, Heinz-Elmar Tenorth, Horst Weishaupt und Manfred Weiß, 2003: *Bildungsberichterstattung Deutschland: Konzeption*. Frankfurt.  
URL: <http://www.kmk.org/doc/publ/bildungsbericht/Konzeption.pdf>.
- Babka von Gostomski, Christian, 1995: Zur Konsistenz und Übereinstimmung von Ehepartnern bei retrospektiv erhobenen Angaben zur Person und zur Beziehung. *Zeitschrift für Familienforschung* 7(1): 6-26.
- Baker, Stuart G., 1994: Missing Data. *Composite Linear Models for Incomplete Multinomial Data*. *Statistics in Medicine* 13: 609-622.
- Baker, Stuart G., und Nan M. Laird, 1988: Regression Analysis for Categorical Variables With Outcome Subject to Nonignorable Nonresponse. *Journal of the American Statistical Association* 83(401): 62-69.
- Baker, Stuart G., William F. Rosenberger und Rebecca DerSimonian, 1992: Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* 11(5): 643-657.
- Basic, Edin, Ivo Marek und Ulrich Rendtel, 2005: The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data. *Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“*, Arbeitspapier Nr. 3. Berlin: Freie Universität.  
URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier3,property=file.pdf>; 29. 06. 2007.
- Basic, Edin, und Ulrich Rendtel, 2005: Estimation strategies in the presence of non-coverage in the German Microcensus-Panel: An evaluation using SOEP data. *Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“*, Arbeitspapier Nr. 8. Berlin: Freie Universität.  
URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier8,property=file.pdf>; 29. 06. 2007.
- Bassi, Francesca, Jacques A. Hagenaars, Marcel A. Croon und Jeroen K. Vermunt, 2000: Estimating True Changes when Categorical Panel Data are Affected by Uncorrelated and Correlated Classification Errors. *Sociological Methods & Research* 29(2): 230-268.
- Baumert, Jürgen, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Petra Stanat, Klaus-Jürgen Tillmann und Manfred Weiß (Hg.), 2001: *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, Jürgen, und Gundel Schümer, 2001: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. S. 323- 407 in: Jürgen Baumert, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Petra Stanat, Klaus-Jürgen Tillmann

- und Manfred Weiß (Hg.): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Beicht, Ursula, Klaus Troltsch, Günter Walden und Rudolf Werner, 2003: Technische Berufe im dualen System der Berufsausbildung - Stellenwert und Entwicklungstendenzen. Gutachten im Rahmen der Berichterstattung zur technologischen Leistungsfähigkeit Deutschlands. Studien zum deutschen Innovationssystem, Nr. 3-2003. Bonn: Bundesinstitut für Berufsbildung.
- Bellenberg, Gabriele, 1999: Individuelle Schullaufbahnen. Eine empirische Untersuchung über Bildungsverläufe von der Einschulung bis zum Abschluß. Weinheim: Juventa.
- Biemer, Paul M., 2004: An Analysis of Classification Error for the Revised Current Population Survey Employment Questions. *Survey Methodology* 30(2): 127-140.
- Biemer, Paul M., und Lars E. Lyberg, 2003: *Introduction to Survey Quality*. Hoboken, NJ: Wiley.
- Bispinck, Reinhard, Marlies Dorsch-Schweizer und Johannes Kirsch, 2002: Tarifliche Ausbildungsförderung begrenzt erfolgreich – eine empirische Wirkungsanalyse. *WSI Mitteilungen* (4): 213-219.
- Black, Dan, Seth Sanders und Lowell Taylor, 2003: Measurement of Higher Education in the Census and CPS. *Journal of the American Statistical Association* 98(463): 545-554.
- Blossfeld, Hans-Peter, Alfred Hamerle und Karl Ulrich Mayer: 1986: Ereignisanalyse. Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften. Frankfurt: Campus.
- Blossfeld, Hans-Peter, und Gerald Prein (Hg.), 1998: *Rational Choice Theory and Large-Scale Data Analysis*. Boulder, CO: Westview.
- Blossfeld, Hans-Peter, und Yossi Shavit, 1993: Dauerhafte Ungleichheiten. Zur Veränderung des Einflusses der sozialen Herkunft auf die Bildungschancen in dreizehn industrialisierten Ländern. *Zeitschrift für Pädagogik* 39(1): 25-52.
- BMA [Bundesministerium für Arbeit und Sozialordnung], 2001: Lebenslagen in Deutschland. Der erste Armuts- und Reichtumsbericht der Bundesregierung. Bonn: BMA.
- BMBF [Bundesministerium für Bildung und Forschung], 2005: Berufsbildungsbericht 2005. Bonn. URL: [www.bmbf.de/pub/bbb\\_2005.pdf](http://www.bmbf.de/pub/bbb_2005.pdf).
- BMBF, 2003: Berufsbildungsbericht 2003. Bonn. URL: [www.bmbf.de/pub/bbb2003.pdf](http://www.bmbf.de/pub/bbb2003.pdf).
- BMBF, 2002: Grund und Strukturdaten 2001/2002. Bonn. URL: [www.kmu-info.bmbf.de/pdf/GuS2002\\_ges\\_dt.pdf](http://www.kmu-info.bmbf.de/pdf/GuS2002_ges_dt.pdf).
- BMBF, 2001: Berufsbildungsbericht 2001. Bonn. URL: [www.bmbf.de/pub/bbb2001.pdf](http://www.bmbf.de/pub/bbb2001.pdf).
- BMBF, 1999: Berufsbildungsbericht 1999. Bonn. URL: [www.bmbf.de/pub/bbb1999.pdf](http://www.bmbf.de/pub/bbb1999.pdf).
- Böltken, Ferdinand, 1997: Neue siedlungsstrukturelle Gemeindetypen für die Raumbearbeitung. *Mitteilungen und Informationen der BfLR* 3: 4-5.
- Böltken, Ferdinand, und Eleonore Irmen 1997: Neue siedlungsstrukturelle Regions- und Kreistypen. *Mitteilungen und Informationen der BfLR* 1: 4-5.
- Boudon, Raymond, 1974: *Education, Opportunity, and Social Inequality. Changing Prospects in Western Society*. New York: Wiley.
- Bourdieu, Pierre, 1973: Kulturelle Reproduktion und soziale Reproduktion. S. 88-137 [Teil 2] in: Ders.: *Grundlagen einer Theorie der symbolischen Gewalt*. Frankfurt: Suhrkamp.
- Bourdieu, Pierre, 1982: *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt: Campus.
- Bourdieu, Pierre, und Jean-Claude Passeron, 1971: *Die Illusion der Chancengleichheit*. Stuttgart: Klett.



- Brauns, Hildegard, Markus Gangl und Stefani Scherer, 2000: The Educational Stratification of Unemployment Risks at the Beginning of Working Life. Vortrag zur 2. Nutzerkonferenz „Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt“, Mannheim, 12./13. Oktober 2000.  
URL: [www.gesis.org/Dauerbeobachtung/GMLService/Veranstaltungen/NK 2000/paper/text\\_brauns.pdf](http://www.gesis.org/Dauerbeobachtung/GMLService/Veranstaltungen/NK_2000/paper/text_brauns.pdf).
- Brauns, Hildegard, und Susanne Steinmann, 1999: Educational Reform in France, West-Germany and the United Kingdom. *ZUMA-Nachrichten* 44: 7-44.
- Breen, Richard, und John H. Goldthorpe, 1997: Explaining Educational Differentials. Towards a Formal Rational Choice Theory. *Rationality and Society* 9(3): 275-305.
- Breiholz, Holger, 2000: Ergebnisse des Mikrozensus 1999. *Wirtschaft und Statistik* (5): 328-336.
- Brick, J. Michael, Jacque Wernimont und Maida Montes, 1996: The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component. NCES Working Paper No. 96-14. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Büchel, Felix, und Gernot Weißhuhn, 1995: Bildungswege und Berufseintritt im Wandel: mittelfristige Entwicklung und sozio-ökonomische Bestimmungsfaktoren der Bildungsnachfrage und der Übergangsmuster zwischen beruflichen Ausbildungsformen in Deutschland. Bielefeld: Bertelsmann.
- Bundesanstalt für Arbeit, 1973: Verzeichnis der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit. Nürnberg. Bundesanstalt für Arbeit.
- Bundesanstalt für Arbeit, 1988: Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Gliederung nach Berufsklassen für die Statistik der Bundesanstalt für Arbeit. Nürnberg. Bundesanstalt für Arbeit.
- Chambers, Ray L., und Alan H. Welsh 1993: Log-linear models for survey data with non-ignorable non-response. *Journal of the Royal Statistical Society (B)* 55: 151-170.
- Clarke, Paul S., und Pamela F. Tate, 1999: Methodological issues in the production and analysis of longitudinal data from the Labour Force Survey. *National Statistics Methodology Series* 14. London: Government Statistical Service.
- Conaway, Mark R., 1992: The Analysis of Repeated Categorical Measurements Subject to Non-ignorable Nonresponse. *Journal of the American Statistical Association* 87(419): 817-824.
- Conaway, Mark R., 1993: Non-ignorable Non-response Models for Time-ordered Categorical Variables. *Applied Statistics* 42(1): 105-115.
- Copeland, Kennon R., 2004: Panel survey estimation in the presence of late reporting and nonresponse. Dissertation. College Park, MD: University of Maryland.  
URL: <https://drum.umd.edu/dspace/bitstream/1903/1762/1/umi-umd-1740.pdf>.
- Cortina, Kai S., und Luitgard Trommer, 2003: Bildungswege und Bildungsbiographien in der Sekundarstufe I. S. 342-391 in: Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ulrich Mayer und Luitgard Trommer (Hg.): *Das Bildungswesen in der Bundesrepublik Deutschland*. Hamburg: Rowohlt.
- Davidov, Eldad, Peter Schmidt und Sebastian Bamberg, 2002: Time and Money: An Empirical Explanation of Behaviour in the Context of Travel-Mode Choice with the German Microcensus. *European Sociological Review* 19(3): 267-280.
- Dawe, Fiona, und Ian Knight, 1997: A study of proxy response on the Labour Force Survey. *Survey Methodology Bulletin* 40: 30-36.

- Dempster, Arthur P., Nan M. Laird und Donald B. Rubin, 1977: Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society (B)* 39: 1-38.
- Diekmann, Andreas, und Thomas Voss, 2003: Die Theorie rationalen Handelns. Stand und Perspektiven. S. 13-29 in: diess. (Hg.): *Rational Choice Theorie in den Sozialwissenschaften. Probleme und Perspektiven*. München: Oldenbourg.
- Dietrich, Hans, und Lutz Bellmann, 1999: Das Ausbildungsverhalten der bundesdeutschen Betriebe. S. 75-104 in: Andreas Flitner, Christian Petry und Ingo Richter (Hg.): *Wege aus der Ausbildungskrise*. Opladen: Leske+Budrich.
- Dräther, Hendrik, Uwe Fachinger und Angelika Oelschläger, 2001: Selbständige und ihre Altersvorsorge – Möglichkeiten der Analyse anhand der Mikrozensus und erste Ergebnisse. ZeS-Arbeitspapier Nr. 1/01. Universität Bremen: Zentrum für Sozialpolitik.
- Duncan, Greg, 2000: Using panel studies to understand household behaviour and well-being. S. 54-75 in: David Rose (Hg.): *Researching Social and Economic Change*. London: Routledge.
- Eggen, Bernd, 1999: Wandel familialer und ökonomischer Lebenslagen junger Erwachsener. S. 188-211 in: Paul Lüttinger (Hg.): *Sozialstrukturanalyse mit dem Mikrozensus*. ZUMA-Nachrichten Spezial Band 6. Mannheim: ZUMA.
- Elias, Peter, 1997: Occupational Classification: concepts, methods, reliability, validity and cross-national comparability. LES Working Paper. Luxembourg.  
URL: [www.lisproject.org/publications/leswps/leswp5.pdf](http://www.lisproject.org/publications/leswps/leswp5.pdf).
- Emmerling, Dieter, und Thomas Riede, 1997: 40 Jahre Mikrozensus. *Wirtschaft und Statistik* (3): 160-174.
- Erikson, Robert, und Jan O. Jonsson, 1996: The Swedish Context: Educational Reform and Long-term Change in Educational Inequality. S. 65-93 in: diess. (Hg.): *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder: Westview Press.
- Ermisch, John, 1996: Analysis of Leaving the Parental Home and Returning to it Using Panel Data. Working papers of the ESRC Research Centre on Micro-social change, Paper 96-1. Colchester: University of Essex.
- Esser, Hartmut, 1999: *Soziologie. Spezielle Grundlagen*. Bd. 1. Situationslogik und Handeln. Frankfurt: Campus.
- Esser, Hartmut, Heinz Grohmann, Walter Müller und Karl-August Schäffer, 1989: Mikrozensus im Wandel. Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung. In: Statistisches Bundesamt (Hg.): *Forum der Bundesstatistik*, Band 11. Stuttgart: Metzler-Poeschel.
- Euwals, Rob, und Rainer Winkelmann, 2001: Why Do Firms Train? Empirical Evidence on the First Labour Market Outcomes of Graduated Apprentices. IZA Discussion Paper No. 319. Bonn: IZA. URL: [ftp://ftp.iza.org/dps/dp319.pdf](http://ftp.iza.org/dps/dp319.pdf).
- Fay, Robert E., 1986: Causal Models for Patterns of Nonresponse. *Journal of the American Statistical Association* 81(394): 354-365.
- Fitzenberger, Bernd, und Alexandra Spitz, 2004: Die Anatomie des Berufswechsels: Eine empirische Bestandsaufnahme auf Basis der BIBB/IAB-Daten 1998/99. ZEW Discussion Paper No. 04-05. Mannheim: ZEW. URL: [ftp://ftp.zew.de/pub/zew-docs/dp/dp0405.pdf](http://ftp.zew.de/pub/zew-docs/dp/dp0405.pdf).
- Fitzmaurice, Garrett M., Peter Clifford und Anthony F. Heath, 1996: Logistic Regression Models for Binary Panel Data with Attrition. *Journal of the Royal Statistical Society A* 159(2): 249-263.

- Forster, Jonathan J., und Peter W.F. Smith, 1998: Model-based inference for categorical survey data subject to non-ignorable non-response. *Journal of the Royal Statistical Society B* 60(1): 57-70.
- Franz, Wolfgang, 1999: Arbeitsmarktökonomik. 4., überarb. Auflage. Berlin: Springer.
- Franz, Wolfgang, Joachim Inkmann, Winfried Pohlmeier und Volker Zimmermann, 2000: Young and Out in Germany. On Youth's Chances of Labour Market Entrance in Germany. S. 381-425 in: David G. Blanchflower und Richard B. Freeman (Hg.): *Youth Employment and Joblessness in Advanced Countries*. Chicago: University Press.
- Frick, Joachim, 1996: Lebenslagen im Wandel: Determinanten kleinräumlicher Mobilität in Westdeutschland. In: Projektgruppe „Das Sozio-ökonomische Panel“ im Deutschen Institut für Wirtschaftsforschung, Berlin, (Hg.): *Sozio-ökonomische Daten und Analysen für die Bundesrepublik Deutschland*; Band 9. Frankfurt: Campus.
- Gangl, Markus, und Walter Müller, 1999: Germany. S. 50-95 in: Damian F. Hannan, Emer Smyth, Selina McCoy et al.: *A Comparative Analysis of Transitions from Education to Work in Europe (CATEWE)*. Volume 2: Country Reports - France, Germany, Ireland, The Netherlands, Scotland, Portugal. ESRI Working Paper No. 118(b). Dublin: Economic and Social Research Institute.
- URL: <http://www.mzes.uni-mannheim.de/projekte/catewe/papers/vol2.zip>.
- Gartner, Hermann, 2000: Das Auszugsverhalten junger Menschen aus dem Elternhaus in Westdeutschland seit 1984. Discussionpaper 215. München: Sfb 386.
- Georg, Werner, Christiane Strzoda und Jürgen Zinnecker, 1994: Determinanten des Auszugs junger Erwachsener aus dem Elternhaus. Eine Analyse mit Survivalmodellen. *ZA-Information* 34: 106-123.
- Glynn, Robert J., Nan M. Laird und Donald B. Rubin, 1986: Selection modeling versus mixture modeling with non-ignorable nonresponse. S. 115-152 in Howard Wainer: *Drawing Inferences from Self-selected Samples*. New York: Springer.
- Goodman, Leo A., 1973: The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika* 60: 179-192.
- Granato, Nadia, 2000: CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. Mikrodaten-Tools. Mannheim: ZUMA.
- URL: [www.gesis.org/Dauerbeobachtung/GML/service/Mikrodaten-Tools/casmin5.pdf](http://www.gesis.org/Dauerbeobachtung/GML/service/Mikrodaten-Tools/casmin5.pdf)
- Granato, Nadia, und Frank Kalter, 2001: Die Persistenz ethnischer Ungleichheit auf dem deutschen Arbeitsmarkt: Diskriminierung oder Unterinvestition in Humankapital? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53(3): 497-520.
- Guo, Guang, 1993: Event-History for Left-Truncated Data. S. 217-243 in: Peter V. Marsden (Hg.): *Sociological Methodology 1993*; Vol. 23. Oxford: Blackwell.
- Haas, Anette, 2002: Arbeitsplatzmobilität nach Abschluss einer dualen Berufsausbildung. Befunde aus der IAB-Historikdatei 1992 bis 1997. IAB Werkstattbericht Nr. 3. Nürnberg: IAB.
- Hall, Anja, (o.J.) [2004]: Arbeitsmarkterfolg von Absolventen des dualen Systems und von Berufsfachschulabsolventen im Vergleich – Analysen auf der BIBB/IAB-Erhebung 1998/99. Bonn: BIBB.
- URL: [www.bibb.de/dokumente/pdf/DS-BFS\\_BIBBIAB.pdf](http://www.bibb.de/dokumente/pdf/DS-BFS_BIBBIAB.pdf).
- Hamann, Silke, Gerhard Krug, Markus Köhler, Wolfgang Ludwig-Mayerhofer und Anne Hackett, 2004: Die IAB-Regionalstichprobe 1975-2001: IABS-R01. *ZA-Information* 55: 34-59.
- Handl, Johann, 1996: Hat sich die berufliche Wertigkeit der Bildungsabschlüsse in den achtziger Jahren verringert? Eine Analyse der abhängigen erwerbstätigen, deutschen Berufsanfänger auf

- der Basis von Mikrozensusergebnissen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48(2): 249-273.
- Haslinger, Alois, und Josef Kytir, 2006: Stichprobendesign, Stichprobenziehung und Hochrechnung des Mikrozensus ab 2004. *Statistische Nachrichten* 61(6): 510-519.
- Heckman, James J., 1976: The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimation method for such models. *Annals of Economic and Social Measurement* 5: 475-492.
- Heidenreich, Hans-Joachim, 1994: Hochrechnung des Mikrozensus ab 1990. S. 112-123 in: Siegfried Gabler, Jürgen H. P. Hoffmeyer-Zlotnik und Dagmar Krebs (Hg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Heidenreich, Hans-Joachim, 2002: Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten. *Allgemeines Statistisches Archiv* 86(2): 213-231.
- Heidenreich, Hans-Joachim, und Michaela Nöthen, 2002: Der Wandel der Lebensformen im Spiegel des Mikrozensus. *Wirtschaft und Statistik*, 1: 26-38.
- Hensher, David A., und Lester W. Johnson, 1981: *Applied Discrete-Choice Modelling*. London: Croom Helm/New York: Wiley.
- Henz, Ursula, und Ineke Maas, 1995: Chancengleichheit durch die Bildungsexpansion? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47(4): 605-633.
- Herberger, Lothar, 1973: *Praktische Erfahrungen mit Verlaufsstatistiken*. *Allgemeines Statistisches Archiv* 57: 54-76.
- Herter-Eschweiler, Robert, 2003: Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. Bonn: Statistisches Bundesamt (unveröffentlichtes Manuskript).
- Hillman, Kylie J., und Gary N. Marks, 2002: *Becoming an Adult: Leaving home, relationships and home ownership among Australian youth*. Longitudinal Survey of Australian Youth (LSAY) Research Report Number 28. Melbourne: Australian Council for Educational Research.
- Hillmert, Steffen, und Marita Jacob, 2005: Zur Entwicklung herkunftsbezogener Ungleichheiten im Bildungsverlauf. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 57(3): 414-442.
- Huinink, Johannes, und Dirk Konietzka, 2000: Leaving Parental Home in the Federal Republic of Germany and the GDR. The changing interrelations of leaving home and other transition events to adulthood. Paper presented on the Workshop on "Leaving Home - A European Focus", 6.-8. September 2000. Max Planck Institute for Demographic Research, Rostock. URL: [www.demogr.mpg.de/Papers/workshop000906\\_paper11.pdf](http://www.demogr.mpg.de/Papers/workshop000906_paper11.pdf); 01.08.2001.
- Iacovou, Maria, 2001: Leaving Home in the European Union. Working Papers of the Institute for Social and Economic Research, paper 2001-18. Colchester: University of Essex.
- Juang, Linda P., Rainer K. Silbereisen und Margit Wiesner, 1999: Predictors of leaving home in young adults raised in Germany: A Replication of a 1991 study. *Journal of Marriage & Family* 61: 505-515.
- Kalter, Frank, 1997: Wohnortwechsel in Deutschland. Ein Beitrag zur Migrationstheorie und zur empirischen Anwendung von Rational-Choice-Modellen. Opladen: Leske + Budrich.
- Kalter, Frank, 2000: Theorien der Migration. S. 438-457 in: Ulrich Mueller, Bernhard Nauck und Andreas Diekmann (Hg.): *Handbuch der Demographie*, 1. Modelle und Methoden. Berlin: Springer.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland], 2001: *Grundstruktur des Bildungswesens in der Bundesrepublik Deutschland*.

- Diagramm 1999 für Faltblatt, Sonderdruck und Download. Bonn.  
URL: [www.kmk.org/schul/home.htm](http://www.kmk.org/schul/home.htm) - .../doku/ddiagr.doc.
- KMK, 2003: Das Bildungswesen in der Bundesrepublik Deutschland 2002. Bonn.  
URL: [www.kmk.org/dossier/dossier\\_2002/dossier\\_2002\\_dt\\_ebook.pdf](http://www.kmk.org/dossier/dossier_2002/dossier_2002_dt_ebook.pdf).
- Körner, Thomas, und Jürgen Schmidt, 2006: Qualitätsberichte - ein neues Informationsangebot über Methoden, Definitionen und Datenqualität der Bundesstatistiken. *Wirtschaft und Statistik* (2): 109-117.
- Koller, Siegfried, und Lothar Herberger, 1960: Der Mikrozensus. *Allgemeines Statistisches Archiv* 44(3): 205-254.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hg.), 2001: Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden: Nomos.
- Konietzka, Dirk, 1999: Die Verberuflichung von Marktchancen. Die Bedeutung des Ausbildungsberufs für die Plazierung im Arbeitsmarkt. *Zeitschrift für Soziologie* 28 (5): 379-400.
- Konietzka, Dirk, 2004: Soziale Ungleichheiten in der beruflichen Ausbildung und beim Übergang in den Arbeitsmarkt. S. 281-309 in: Rolf Becker und Wolfgang Lauterbach (Hg.): *Bildung als Privileg? Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Konietzka, Dirk, und Johannes Huinink, 2003: Die De-Standardisierung einer Statuspassage? Zum Wandel des Auszugs aus dem Elternhaus und des Übergangs in das Erwachsenenalter in Westdeutschland. *Soziale Welt* 54(3): 285-312.
- Konold, Michael 2005: Der Ausfall räumlich mobiler Personen im Mikrozensus-Panel und seine Auswirkungen auf Übergangsanalysen im Familien- und Erwerbsbereich. Ergebnisse empirischer Untersuchungen. Düsseldorf: Landesamt für Datenverarbeitung und Statistik NRW (unveröff. Manuskript).
- Korn, Edward L., Barry I. Graubard und Douglas Midthune, 1997: Time-to-Event-Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. *American Journal of Epidemiology* 145(1): 72-80.
- Kristen, Cornelia, 1999: Bildungsentscheidungen und Bildungsungleichheit – ein Überblick über den Forschungsstand. Arbeitspapiere Nr. 5. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Kristen, Cornelia, Anika Römmers, Walter Müller und Frank Kalter, 2005: Längsschnittstudien für die Bildungsberichterstattung – Beispiele aus Europa und Nordamerika. In: Bundesministerium für Bildung und Forschung (Hg.). *Bildungsreform*, Band 10. Berlin.  
URL: [http://www.bmbf.de/pub/Laengsschnitt\\_fuer\\_Bildungsberichterstattung.pdf](http://www.bmbf.de/pub/Laengsschnitt_fuer_Bildungsberichterstattung.pdf).
- Laaksonen, Seppo, und Ray Chambers, 2006: Survey Estimation Under Informative Nonresponse with Follow-up. *Journal of Official Statistics* 22(1): 81-95.
- Lauterbach, Wolfgang, und Kurt Lüscher, 1999: Wer sind die Spätauszieger? Oder: Herkunftsfamilie, Wohnumfeld und die Gründung eines eigenen Haushalts. Arbeitspapier Nr. 33. Konstanz: Sozialwissenschaftliche Fakultät.
- Lehmann, Thomas, 2004: Behandlung von fehlenden Werten bei nicht ignorierbaren Ausfallmechanismen. Dissertation zur Erlangung des akademischen Grades doctor rerum politicarum, Friedrich-Schiller-Universität Jena, Wirtschaftswissenschaftliche Fakultät, 19.10.2004.

- Lengerer, Andrea, Jeanette Bohr und Andrea Janßen, 2005: Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. ZUMA-Arbeitsbericht 2005/05. Mannheim: ZUMA.
- Little, Roderick J.A., 1985: Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data. *Bulletin of the International Statistical Institute* 15: 1-15.
- Little, Roderick J. A., 1993: Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88: 125–134.
- Little, Roderick J.A., und Donald B. Rubin, 2002: *Statistical Analysis with Missing Data*. 2. Auflage. New York: Wiley.
- Lois, Daniel, 2005: Weiterbildungsbeteiligung älterer Erwerbstätiger – Die Messung im Mikrozensus und der Einfluss soziodemografischer Variablen. Beitrag zur 4. Nutzerkonferenz „Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt“, Mannheim, 12./13.10.2005.  
URL: [www.gesis.org/Dauerbeobachtung/GML/Service/Veranstaltungen/4.NK\\_2005/papers/18\\_Lois.pdf](http://www.gesis.org/Dauerbeobachtung/GML/Service/Veranstaltungen/4.NK_2005/papers/18_Lois.pdf).
- Lotze, Sabine, und Holger Breiholz, 2002a: Zum neuen Erhebungsdesign des Mikrozensus. Teil 1. *Wirtschaft und Statistik*, 5: 359-366.
- Lotze, Sabine, und Holger Breiholz, 2002b: Zum neuen Erhebungsdesign des Mikrozensus. Teil 2. *Wirtschaft und Statistik*, 6: 454-459.
- Lüttinger, Paul und Thomas Riede, 1997: Der Mikrozensus. Amtliche Daten für die Sozialforschung. ZUMA-Nachrichten Nr. 41: 19-43.
- Madrian, Brigitte, und Lars John Lefgren, 2000: An Approach To Longitudinally Matching Population Survey (CPS) Respondents. *Journal of Economic and Social Measurement* 26(1): 31-62.
- Mare, Robert D., 1980: Social Background and School Continuation Decisions. *Journal of the American Statistical Association* 75(370): 295-305.
- Mare, Robert D., 1981: Change and Stability in Educational Stratification. *American Sociological Review* 46(February): 72-87.
- Martin, Jean, John Bynner, Graham Kalton, Paul Boyle, Harvey Goldstein, Vernon Gayle, Samantha Parsons und Andrea Piesse, 2006: Strategic Review of Panel and Cohort Studies. Report to the Research Resources Board of the Economic and Social Research Council. London: Longview.  
URL: <http://www.longviewuk.com/pages/documents/ESRCstrategicreviewwebsite2bedit27.03.06a.pdf>.
- Mayer, Karl Ulrich, und Michael Wagner, 1986: Der Auszug von Kindern aus dem elterlichen Haushalt – ein Erklärungsmodell für die Geburtsjahrgänge 1929-31, 1939-41 und 1949-51, S. 1-42 in: Klaus F. Zimmermann (Hg.): *Demographische Probleme der Haushaltsökonomie. Beiträge zur Quantitativen Ökonomie*, Band IX, Bochum: Brockmeyer.
- McGovern, Pamela D., und John M. Bushery, 1999: Data mining the CPS reinterview: Digging into response error. URL: [www.fcs.m.gov/99papers/mcgovern.pdf](http://www.fcs.m.gov/99papers/mcgovern.pdf).
- Meyer, Kurt, 1994: Zum Auswahlplan des Mikrozensus ab 1990. S. 106-111 in: Siegfried Gabler, Jürgen H.P. Hoffmeyer-Zlotnik und Dagmar Krebs (Hg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Molenberghs, Geert, und Geert Verbeke, 2005: *Models for Discrete Longitudinal Data*. New York: Springer.



- Molenberghs, Geert, und Els Goetghebeur, 1997: Simple Fitting Algorithms for Incomplete Categorical Data. *Journal of the Royal Statistical Society B (Methodological)* 59(2): 401-414.
- Molenberghs, Geert, Els J. T. Goetghebeur, Stuart R. Lipsitz und Michael G. Kenward, 1999: Nonrandom Missingness in Categorical Data: Strengths and Limitations. *The American Statistician* 53(2): 110-118.
- Müller, Walter: 2005: Education and Youth Integration into European Labour Markets. *International Journal of Comparative Sociology* 46(5-6): 461-485.
- Müller, Walter, 2003: Nutzung vorhandener statistischer Daten. S. 57-67 in: Statistisches Bundesamt (Hg.): *Flexibilisierung der amtlichen Statistik*. Forum der Bundesstatistik, 40. Wiesbaden: Statistisches Bundesamt.
- Müller, Walter, und Dietmar Haun, 1994: Bildungsungleichheit im sozialen Wandel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 46(1): 1-42.
- Mulder, Clara H., William A.V. Clark, und Michael Wagner, 2002: A comparative analysis of leaving home in the United States, the Netherlands and West Germany. *Demographic Research*, Vol. 7, Article 17, S. 565-592.  
URL: [www.demographic-research.org/Volumes/Vol7/17/7-17.pdf](http://www.demographic-research.org/Volumes/Vol7/17/7-17.pdf); 23.05.04.
- Nauck, Bernhard, 2000: Eltern-Kind-Beziehungen in Migrantenfamilien - ein Vergleich zwischen griechischen, italienischen, türkischen und vietnamesischen Familien in Deutschland. S. 347-392 in: Sachverständigenkommission 6. Familienbericht (Hg.): *Familien ausländischer Herkunft in Deutschland*. Empirische Beiträge zur Familienentwicklung und Akkulturation. Materialien zum 6. Familienbericht, Band I. Opladen: Leske + Budrich.
- Nauck, Bernhard, 2001: Solidarpotenziale von Migrantenfamilien. Expertise. Friedrich-Ebert-Stiftung, Gesprächskreis Migration und Integration. Chemnitz.  
URL: <http://fesportal.fes.de/pls/portal30/docs/FOLDER/BERATUNGSZENTRUM/ASPOL/NAUCKEXPERTISE.DOC>.
- NCES [U.S. Department of Education, National Center for Education Statistics], 1997: *Measurement Error Studies at the National Center for Education Statistics*, NCES 97-464. Washington, DC: NCES.
- Nöthen, Manuela, 2005: Von der „traditionellen Familie“ zu „neuen Lebensformen“. Neuerungen in der Familienberichterstattung des Mikrozensus. *Wirtschaft und Statistik* (1): 25-40.
- Office of Management and Budget, 2001: *Measuring and Reporting Sources of Error in Surveys*. Statistical Policy Working Paper 31. URL: [www.fcsm.gov/01papers/SPWP31\\_final.pdf](http://www.fcsm.gov/01papers/SPWP31_final.pdf).
- Opp, Karl-Dieter, 1999: Contending Conceptions of the Theory of Rational Action. *Journal of Theoretical Politics* 11(2): 171-202.
- Park, Taesung, und Morton B. Brown 1994: Models for categorical data with nonignorable nonresponse. *Journal of the American Statistical Association* 89: 44-52.
- Pöschl, Hannelore, 1992: Geringfügige Beschäftigung 1990. Ergebnisse des Mikrozensus. *Wirtschaft und Statistik* (3): 166-170.
- Pötter, Ulrich, 2006: Probabilistische Selektionsmodelle. S. 172-202 in Andreas Diekmann (Hg.): *Methoden der Sozialforschung*. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 44/2004. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Pötzsch, Ina, 1982: *Das Bildungsschicksal von qualifizierten Haupt- und Realschülern in der neugestalteten gymnasialen Oberstufe – ein Vergleich mit grundständigen Gymnasiasten*. Dissertation. Erziehungswissenschaftliche Fakultät der Universität zu Köln.



- Porst, Rolf, und Klaus Zeifang, 1987: A Description of the German General Social Survey Test-Retest Study and a Report on the Stabilities of the Sociodemographic Variables. *Sociological Methods & Research* 15(3): 177-218.
- Rässler, Susanne, 2000: Ergänzung fehlender Daten in Umfragen. *Jahrbücher für Nationalökonomie und Statistik* 200(1): 64-94.
- Reimer, Maike, 2001: Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Verlaufsdaten. Kognitive und erhebungspragmatische Aspekte. Materialien aus der Bildungsforschung Nr. 71. Berlin: Max-Planck-Institut für Bildungsforschung.
- Rendtel, Ulrich, 2005: Wie geeignet ist der Mikrozensus für Längsschnittanalysen? Beitrag für die 4. Nutzerkonferenz „Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt“, Mannheim, 12./13.10.2005.  
URL: [www.gesis.org/Dauerbeobachtung/GML/Service/Veranstaltungen/4.NK\\_2005/papers/rendtel.pdf](http://www.gesis.org/Dauerbeobachtung/GML/Service/Veranstaltungen/4.NK_2005/papers/rendtel.pdf).
- Rendtel, Ulrich, 1995: Lebenslagen im Wandel: Panelfälle und Panelrepräsentativität. Frankfurt: Campus.
- Rendtel, Ulrich, und Bernhard Schimpl-Neimanns, 2001: Die Berechnung der Varianz von Populationsschätzern im Scientific Use File des Mikrozensus ab 1996. *ZUMA-Nachrichten* 48: 85-116.
- Riede, Thomas, und Dieter Emmerling, 1994: Analysen zur Freiwilligkeit der Auskunftserteilung im Mikrozensus. Sind Stichprobenergebnisse bei freiwilliger Auskunftserteilung verzerrt? *Wirtschaft und Statistik* (9): 733-742.
- Rindskopf, David, 1992: A general approach to categorical data analysis with missing data, using generalized linear models with composite links. *Psychometrika* 57(1): 29-42.
- Rogers, William, 1993: sg17: Regression standard errors in clustered samples. *Stata Technical Bulletin* 13:19-23.
- Rohloff, Sandra, 2005: Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelfallstichprobe“, Arbeitspapier Nr. 6. Wiesbaden: Statistisches Bundesamt.  
URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier6,property=file.pdf>; 29. 06. 2007.
- Roloff, Juliane, 2001: Einige Bemerkungen zum Wanderungsgeschehen in Ostdeutschland, insbesondere zwischen Ost- und Westdeutschland, nach ausgewählten Altersgruppen. *BIB-Mitteilungen* 22(2): 14-19.
- Rubin, Donald B., 1976: Inference and Missing Data. *Biometrika* 63(3): 581-592.
- Rudolph, Helmut, 1998: „Geringfügige Beschäftigung“ mit steigender Tendenz. Erhebungskonzepte, Ergebnisse und Interpretationsprobleme der verfügbaren Datenquellen. IAB-Werkstattbericht Nr. 9. Nürnberg: IAB.
- Särndal, Carl-Erik, Bengt Swensson und Jan Wretman, 1997: Model Assisted Survey Sampling. New York: Springer.
- Schachter, Jason, 2001: Why People Move: Exploring the March 2000 Current Population Survey: March 1999 to March 2000. Current Population Reports P23-204. Washington, D.C.: U.S. Census Bureau. URL: [www.census.gov/prod/2001pubs/p23-204.pdf](http://www.census.gov/prod/2001pubs/p23-204.pdf).
- Schaeper, Hildegard, Thomas Kühn und Andreas Witzel, 2000: Diskontinuierliche Erwerbskarrieren und Berufswechsel in den 1990ern: Strukturmuster und biografische Umgangsweisen

- betrieblich ausgebildeter Fachkräfte. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 33(1): 80-100.
- Schafer, Joseph L., 1997: Analysis of Incomplete Multivariate Data. London: Chapman & Hall.
- Schafer, Joseph L., und John W. Graham, 2002: Missing Data: Our View of the State of the Art. Psychological Methods 7(2): 147-177.
- Schimpl-Neimanns, Bernhard, 2000: Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. Kölner Zeitschrift für Soziologie und Sozialpsychologie 52(4): 636-669.
- Schimpl-Neimanns, Bernhard, 2005: Bildungsverläufe im Mikrozensuspanel 1996-1999: Besuch der gymnasialen Oberstufe bis zum Abitur. ZUMA-Arbeitsbericht 2005/02. Mannheim: ZUMA
- Schimpl-Neimanns, Bernhard, 2006a: Filekonzept zum Mikrozensuspanel. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. Arbeitspapier Nr. 12. Mannheim: ZUMA.  
URL: [http://www.gesis.org/dauerbeobachtung/gml/F+E/Publikationen/BSN\\_arbeitspapier12.pdf](http://www.gesis.org/dauerbeobachtung/gml/F+E/Publikationen/BSN_arbeitspapier12.pdf); 29. 06. 2007.
- Schimpl-Neimanns, Bernhard, 2006b: Berufliche Ausbildungsverläufe bis zum Übergang ins Erwerbsleben - Analysen zur Stichprobenselektivität des Mikrozensuspanels 1996-1999. ZUMA-Arbeitsbericht 2006/02. Mannheim: ZUMA.
- Schimpl-Neimanns, Bernhard, 2006c: Zur Datenqualität der Bildungsangaben im Mikrozensus. ZUMA-Arbeitsbericht 2006/03. Mannheim: ZUMA.
- Schimpl-Neimanns, Bernhard, 2006d: Auszug aus dem Elternhaus: Ergebnisse des Mikrozensuspanels 1996-1999. ZUMA-Arbeitsbericht 2006/04. Mannheim: ZUMA.
- Schmidt, Claudia, und Horst Weishaupt, 2004: Projektbericht „Dokumentation der Längsschnittforschung im Bildungsbereich“. Erfurt.  
URL: [http://www.za.uni.koeln.de/data/add\\_studies/kat56/1%E4ngsschnittforschung\\_bildungsbereich/Projektbericht\\_L%E4ngsschnittforschung\\_Bildungsbereich.pdf](http://www.za.uni.koeln.de/data/add_studies/kat56/1%E4ngsschnittforschung_bildungsbereich/Projektbericht_L%E4ngsschnittforschung_Bildungsbereich.pdf).
- Schneider, Thorsten, 2005: Erfolgreich durchs deutsche Schulsystem. Soziale Herkunft und Bildungsbeteiligung. Dissertation. Philosophische Fakultät der Universität Zürich.
- Schöngen, Klaus, 1995: Übernahme, beruflicher Verbleib und berufliche Perspektiven. Westdeutsche Ausbildungsabsolventinnen und -absolventen der Jahre 1990 bis 1993. S. 147-166 in: Gisela Westhoff (Hg.): Übergänge von der Ausbildung in den Beruf. Bielefeld: Bertelsmann.
- Schulz, Florian, und Hans-Peter Blossfeld, 2006: Wie verändert sich die häusliche Arbeitsteilung im Eheverlauf? Eine Längsschnittstudie der ersten 14 Ehejahre in Westdeutschland. Kölner Zeitschrift für Soziologie und Sozialpsychologie 58(1): 23-49.
- Schupp, Jürgen, Roland Habich und Wolfgang Zapf, 1996: Sozialberichterstattung im Längsschnitt – Auf dem Weg zu einer dynamischen Sicht der Wohlfahrtsproduktion. S. 11-45 in: Wolfgang Zapf, Jürgen Schupp und Roland Habich (Hg.): Lebenslagen im Wandel: Sozialberichterstattung im Längsschnitt. Frankfurt: Campus.
- Schupp, Jürgen, Joachim Frick, Lutz Kaiser und Gert Wagner, 1999: Zur Erhebungsproblematik geringfügiger Beschäftigung. Ein Vergleich des Mikrozensus mit dem Sozio-ökonomischen Panel und dem Europäischen Haushaltspanel. S. 93-118 in: Paul Lüttinger, (Hg.): Sozialstrukturanalysen mit dem Mikrozensus. ZUMA-Nachrichten Spezial, Band 6. Mannheim: ZUMA.
- Schwarz, Karl, 1989: Wann verlassen die Kinder das Elternhaus? Zeitschrift für Bevölkerungswissenschaft 15(1): 39-58.

- Schwerdt, Wolfgang, und Stefan Bender, 2003: Was tun Lehrlinge nach ihrer Ausbildung? Eine Analyse mit dem Linked-Employer-Employee-Datensatz des IAB. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 36(1): 46-59.
- Seifert, Wolfgang, und Stefan Bender, 1996: Zuwanderer auf dem Arbeitsmarkt: nationalitäten- und geschlechtsspezifische Unterschiede. *Zeitschrift für Soziologie* 25(6): 473-495.
- Sengenberger, Werner, 1987: Struktur und Funktionsweise von Arbeitsmärkten. Die Bundesrepublik im internationalen Vergleich. Frankfurt: Campus.
- Singer, Judith D., und John B. Willett, 2003: *Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence*. New York: Oxford University Press.
- SIPP, 2001: Survey of Income and Program Participation. User's Guide. In: U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau (Hg.). Washington, D.C. URL: <http://www.bls.census.gov/sipp/usrguide/sipp2001.pdf>.
- Skinner, Chris, 2000: Dealing with measurement error in panel analysis. S. 113-125 in: David Rose (Hg.): *Researching Social and Economic Change*. London: Routledge.
- Skrondal, Anders, und Sophia Rabe-Hesketh, 2004: *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall.
- Statistik Austria, 2004: Sonderprogramme im Rahmen des Mikrozensus 1969 bis 2003. Wien. URL: <http://www.statistik.at/mikrozensus/mikrozensusalt.pdf>.
- Statistische Ämter des Bundes und der Länder, 2004: Ergebnisse des Zensus 2001. *Wirtschaft und Statistik* (8): 813-833.
- Statistisches Bundesamt, 1982: *Systematik der Wirtschaftszweige mit Betriebs- und ähnlichen Benennungen*. Ausgabe 1979. Stuttgart: Kohlhammer.
- Statistisches Bundesamt, 1990: *Fachserie 11 Bildung und Kultur, Reihe 1 Allgemeinbildende Schulen. Schuljahr 1988/89*. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1992: *Fachserie 11 Bildung und Kultur, Reihe 1 Allgemeinbildende Schulen. Schuljahr 1990/91*. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1994: *Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1994: *Klassifikation der Wirtschaftszweige mit Erläuterungen*. Ausgabe 1993. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1996: *Fachserie 11 Bildung und Kultur, Reihe 1 Allgemeinbildende Schulen. Schuljahr 1995/96*. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1997a: *Arbeitsunterlage Allgemeinbildende Schulen Schuljahr 1996-97. Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 1 - Allgemeinbildende Schulen*. Wiesbaden.
- Statistisches Bundesamt, 1997b: *Arbeitsunterlage Berufliche Schulen 1995/96. Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 2 - Berufliche Schulen Schuljahr 1995/96*. Wiesbaden.
- Statistisches Bundesamt, 1998a: *Allgemeinbildende Schulen, Schuljahr 1997/98, Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 1 Allgemeinbildende Schulen, Arbeitsunterlage*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 1998b: *Berufliche Schulen 1996/97, Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 2 Berufliche Schulen Schuljahr 1996/97, Arbeitsunterlage*. Wiesbaden: Statistisches Bundesamt.

- Statistisches Bundesamt, 1999a: Zum Auswahlplan des Mikrozensus ab 1990. S. E2 49-54 in: Statistisches Bundesamt (Hg.): Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957 (Loseblattsammlung). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 1999b: Berufliche Schulen 1998/99. Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 2 Berufliche Schulen Schuljahr 1998/99. Arbeitsunterlage. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 1999c: Allgemeinbildende Schulen, Schuljahr 1998/99, Ergänzende Tabellen zur Fachserie 11 Bildung und Kultur, Reihe 1 Allgemeinbildende Schulen, Arbeitsunterlage. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 2000: Bildung und Kultur. Fachserie 11. Reihe 3.S.1 Berufliche Bildung 1978-1999. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt (Hg.), 2006a: Handbuch Mikrozensus-Panel 1996-1999. Methodenverbund Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe. Version 0.2 - Juli 2006. Bonn: Statistisches Bundesamt, Gruppe VIII C (Mikrozensus).
- Statistisches Bundesamt, 2006b: Mikrozensus und Arbeitskräfteerhebungen. Zur Problematik nicht-stichprobenbedingter Fehler. Wiesbaden.  
URL: [http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Veroeffentlichungen/methodenpapier\\_\\_mikrozensus\\_\\_arbeitskraefteerhebung.property=file.pdf](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Veroeffentlichungen/methodenpapier__mikrozensus__arbeitskraefteerhebung.property=file.pdf); 29. 06. 2007.
- Statistisches Landesamt des Freistaates Sachsen, 2003: Regionalisierte sächsische Wanderungsanalyse: Ergebnisse einer Befragung 2002 bei ehemaligen sächsischen Bürgerinnen und Bürgern, die in der Zeit von Januar 2000 bis Juni 2001 in ein anderes Bundesland gezogen sind. Kamenz: Statistisches Landesamt des Freistaates Sachsen.
- Stauder, Johannes, 2002: Neue Typisierungen von Haushalten und Lebensformen für den Mikrozensus. S. 17-34 in: Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (Hg.): Statistische Analysen und Studien NRW; 5. Düsseldorf: LDS NRW.
- Stauder, Johannes, und Wolfgang Hüning, 2004: Die Messung von Äquivalenzeinkommen und Armutsquoten auf der Basis des Mikrozensus. S. 9-32 in: Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (Hg.): Statistische Analysen und Studien NRW; 13. Düsseldorf: LDS NRW.
- Stegmann, Heinz, und Hermine Kraft, 1983: Vom Ausbildungs- zum Arbeitsvertrag. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 16(3): 235-251.
- Stein, Ulrich, 2004: Wanderungsanalyse für Baden-Württemberg. Statistisches Monatsheft Baden-Württemberg (1): 17-22.
- Steinmann, Susanne, 2000: Bildung, Ausbildung und Arbeitsmarktchancen in Deutschland. Eine Studie zum Wandel der Übergänge von der Schule in das Erwerbsleben. Opladen: Leske + Budrich.
- Stolzenberg, Ross M., und Daniel A. Relles, 1997: Tools for Intuition about Sample Selection Bias and its Correction. American Sociological Review 62(3): 494-507.
- Toutenburg, Helge, Christian Heumann und Thomas Nittner, 2004: Statistische Methoden bei unvollständigen Daten. Discussion Paper 380. München: Ludwig-Maximilians-Universität.  
URL: <http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper380.ps>.
- Troltsch, Klaus, László Alex, Richard von Bardeleben und Joachim G. Ulrich, 1999: Jugendliche ohne Berufsausbildung. Eine BIBB/EMNID Untersuchung. Bonn.  
URL: [www.forum-bildung.de/bib/material/bibb.pdf](http://www.forum-bildung.de/bib/material/bibb.pdf).

- Unicon, 2007: CPS Survey Topics. Santa Monica, CA. URL: [www.unicom.com](http://www.unicom.com).
- U.S. Census Bureau, 2004: Meeting the 21th Century Demographic Data Needs – Implementing the American Community Survey. Report 9: Comparing Social Characteristics With Census 2000. Washington, CD: U.S. Census Bureau.
- Vermunt, Jeroen, 1996: Causal log-linear modeling with latent variables and missing data. S. 35-60 in: Uwe Engel und Jost Reinecke (Hg.): Analysis of Change: Advanced Techniques in Panel Data Analysis. Berlin: de Gruyter.
- Vermunt, Jeroen, 1997a: Log-linear models for event histories. Advanced Quantitative Techniques in the Social Sciences; 8. Thousand Oaks: Sage.
- Vermunt, Jeroen, 1997b: LEM: A general program for the analysis of categorical data. Tilburg University.
- Verordnung (EG) Nr. 577/98 des Rates vom 9.3.1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft.  
URL: [http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/SharedContent/Oeffentlich/AZ/ZD/Rechtsgrundlagen/Statistikbereiche/Arbeitsmarkt/085\\_\\_VOArbeitskraefte.property=file.pdf](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/SharedContent/Oeffentlich/AZ/ZD/Rechtsgrundlagen/Statistikbereiche/Arbeitsmarkt/085__VOArbeitskraefte.property=file.pdf); 29. 06. 2007.
- Wagner, Michael, 1989: Räumliche Mobilität im Lebensverlauf. Stuttgart: Enke.
- Wagner, Michael, und Johannes Huinink, 1991. Neuere Trends beim Auszug aus dem Elternhaus. S. 39-62 in: Günter Butler, Hans-Joachim Hoffmann-Novotny und Gerhard Schmitt-Rink (Hg.): Acta Demographica 1991. Heidelberg: Physica-Verlag.
- Weick, Stefan, 1993: Determinanten des Auszugs aus der elterlichen Wohnung. S. 86-108 in: Andreas Diekmann und Stefan Weick (Hg.): Der Lebenszyklus als sozialer Prozeß. Berlin: Duncker & Humblot.
- Weidacher, Alois, 2000: Lebensformen, Partnerschaft und Familiengründung. Griechische, italienische, türkische und deutsche junge Erwachsene. S. 193-227 in: Sachverständigenkommission 6. Familienbericht (Hg.): Familien ausländischer Herkunft in Deutschland. Empirische Beiträge zur Familienentwicklung und Akkulturation. Materialien zum 6. Familienbericht, Band I. Opladen: Leske + Budrich.
- Weishaupt, Horst, und Detlef Fickermann, 2001: Informationelle Infrastruktur im Bereich Bildung und Kultur. Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. In: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hg.), 2001: Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos. [CD-ROM Beilage zur Buchausgabe]
- Weißhuhn, Gernot, 2001: Gutachten zur Bildung in Deutschland. Hg. v. Bundesministerium für Bildung und Forschung. Bonn: BMBF Publik.  
URL: [www.bmbf.de/pub/gutachten\\_zur\\_bildung\\_in\\_deutschland.pdf](http://www.bmbf.de/pub/gutachten_zur_bildung_in_deutschland.pdf).
- White, Lynn, 1994: Coresidence and Leaving Home: Young Adults and Their Parents. Annual Review of Sociology 20: 81-102.
- Windzio, Michael, 2004: Kann der regionale Kontext zur „Arbeitslosenfalle“ werden? Der Einfluss der Arbeitslosigkeit auf die Mobilität zwischen regionalen Arbeitsmärkten in Westdeutschland. Kölner Zeitschrift für Soziologie und Sozialpsychologie 56(2): 257-278.
- Winship, Christopher, Robert D. Mare und John R. Warren, 2002: Latent Class Models for Contingency Tables with Missing Data. S. 408-432 in: Jacques A. Hagenaars und Allan L. McCutcheon (Hg.): Applied Latent Class Analysis. Cambridge, NY: Cambridge University Press.

- Wirth, Heike, 2006: Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific Use File. Arbeitspapier Nr. 11 des Methodenverbunds „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. Mannheim: ZUMA.  
URL: [http://www.gesis.org/dauerbeobachtung/gml/F+E/Publikationen/HW\\_arbeitspapier11.pdf](http://www.gesis.org/dauerbeobachtung/gml/F+E/Publikationen/HW_arbeitspapier11.pdf); 29. 06. 2007.
- Wolbers, Maarten, 2003: Job Mismatches and their Labour Market Effects among School Leavers in Europe. S. 89-120 in: Irena Kogan und Walter Müller (Hg.): School-to-Work Transitions in Europe: Analyses of the EU LFS 2000 Ad Hoc Module. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Wooldridge, Jeffrey M., 2002: Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.
- Ziegler, Rolf, und Diana Schladt, 1993: Auszug aus dem Elternhaus und Hausstandsgründung. S. 66-85 in: Andreas Diekmann und Stefan Weick (Hg.): Der Lebenszyklus als sozialer Prozeß. Berlin: Duncker & Humblot.





# Anhang

## LEM-Programme zu den Selektionsmodellen in Kapitel 3

### 1 MCAR-Programm zum Beispiel in Abschnitt 3.3.1, Tabelle 3.2

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 2	* Anzahl manifester Variablen (X, Y)
dim 2 3 3	* Anzahl Ausprägungen der Variablen R, X, Y
lab R X Y	* Variablenlabel
sub XY X	* Subgruppen mit Angaben zu XY bzw. nur zu X
mod Y X {XY} R XY {R}	* Logit-Modell zum Ausfalltyp MCAR
* mod {XY R}	* Alternative Spezifikation als log-lineares Modell
rec 12	* Anzahl der Datenzeilen in „tab3_1.dat“
rco	* Fallzähler (freq) in „tab3_1.dat“
dat tab3_1.dat	* Datentabelle „tab3_1.dat“, s. u.
dum 1 1 1	* Referenzkategorien der Variablen R, X, Y
wma RXY tab3_2.fre	* Ausgabe geschätzter Zellenbesetzungen

Datentabelle „tab3\_1.dat“ (siehe Tabelle 3.1, Mikrozensuspanel 1996-1999)

X	Y	freq	Beschäftigungsdauer 1996 in Monaten (X)	Ausbildungsstatus im April 1997 (Y)	Subgruppe
1	1	420	* <=12	Azubi	XY
1	2	19	* <=12	Abschluss	
1	3	54	* <=12	Abbruch	
2	1	282	* 13-24	Azubi	
2	2	94	* 13-24	Abschluss	
2	3	43	* 13-24	Abbruch	
3	1	25	* >=25	Azubi	
3	2	167	* >=25	Abschluss	
3	3	19	* >=25	Abbruch	
1	0	92	* <=12	Missing	X
2	0	85	* 13-24	Missing	
3	0	81	* >=25	Missing	

### 2 MAR-Programm zum Beispiel in Abschnitt 3.3.2, Tabelle 3.3

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 2	* Anzahl manifester Variablen (X, Y)
dim 2 3 3	* Anzahl Ausprägungen der Variablen R, X, Y
lab R X Y	* Variablenlabel
sub XY X	* Subgruppen mit Angaben zu XY bzw. nur zu X
mod Y X {XY} R XY {RX}	* Logit-Modell zum Ausfalltyp MAR
* mod {XY RX}	* Alternative Spezifikation als log-lineares Modell
rec 12	* Anzahl der Datenzeilen in „tab3_1.dat“
rco	* Fallzähler (freq) in „tab3_1.dat“
dat tab3_1.dat	* Datentabelle „tab3_1.dat“, s. o.
dum 1 1 1	* Referenzkategorien der Variablen R, X, Y
wma RXY tab3_3.fre	* Ausgabe geschätzter Zellenbesetzungen

### 3 NINR-Programm zum Beispiel in Abschnitt 3.3.3, Tabelle 3.4

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 2	* Anzahl manifester Variablen (X, Y)
dim 2 3 3	* Anzahl Ausprägungen der Variablen R, X, Y
lab R X Y	* Variablenlabel
sub XY X	* Subgruppen mit Angaben zu XY bzw. nur zu X
mod Y X {XY} R XY {RY}	* Logit-Modell zum Ausfalltyp NINR
* mod {XY RY}	* Alternative Spezifikation als log-lineares Modell
rec 12	* Anzahl der Datenzeilen in „tab3_1.dat“
rco	* Fallzähler (freq) in „tab3_1.dat“
dat tab3_1.dat	* Datentabelle „tab3_1.dat“, s. o.
dum 1 1 1	* Referenzkategorien der Variablen R, X, Y
wma RXY tab3_4.fre	* Ausgabe geschätzter Zellenbesetzungen

### 4 MAR-Programm zum Beispiel in Abschnitt 3.3.4, Tabelle 3.6

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 3	* Anzahl manifester Variablen (D, X, Y)
dim 2 2 3 3	* Anzahl Ausprägungen der Variablen R, D, X, Y
lab R D X Y	* Variablenlabel
sub DXY DX	* Subgruppen mit Angaben zu DXY bzw. DX
mod Y DX {XY}	* Logit-Modell zum Ausfalltyp MAR
R DXY {cov(RD,1), wei(RD), cov(RDX,2), wei(RDX)}	
rec 21	* Anzahl der Datenzeilen in „tab3_5.dat“
rco	* Fallzähler (freq) in „tab3_5.dat“
dat tab3_5.dat	* Datentabelle „tab3_5.dat“, s. u.
des [ 0 0 0 1	* Designmatrix für Effekte: cov(RD,1)
0 0 0 0 0 0 0 0 0 1 0	* cov(RDX,2): R=2, D=2, X=2
0 0 0 0 0 0 0 0 0 0 1]	* cov(RDX,2): R=2, D=2, X=3
sta wei(RD) [ 1 1 0 1]	* Startwerte für Effekte RD und RDX:
sta wei(RDX)	* Gewichtung struktureller Nullzellen in den IABS-Daten
[ 1 1 1 1 1 1 0 0 0 1 1 1]	
dum 1 1 1 1	* Referenzkategorien der Variablen R, D, X, Y
wma RDXY tab3_6.fre	* Ausgabe geschätzter Zellenbesetzungen

Datentabelle „tab3\_5.dat“ (siehe Tabelle 3.5,  
Beschäftigtenstichprobe (IABS-R01) ( $D=1$ ), Mikrozensuspanel 1996-1999 ( $D=2$ ))

D	X	Y	freq	Beschäftigungsdauer		Ausbildungsstatus	Subgruppe
				1996 in Monaten (X)		im April 1997 (Y)	
1	1	1	8710	*	$\leq 12$	Azubi	DXY
1	1	2	399	*	$\leq 12$	Abschluss	
1	1	3	841	*	$\leq 12$	Abbruch	
1	2	1	5453	*	13-24	Azubi	
1	2	2	1377	*	13-24	Abschluss	
1	2	3	876	*	13-24	Abbruch	
1	3	1	668	*	$\geq 25$	Azubi	
1	3	2	3319	*	$\geq 25$	Abschluss	
1	3	3	1427	*	$\geq 25$	Abbruch	
2	1	1	420	*	$\leq 12$	Azubi	DX
2	1	2	19	*	$\leq 12$	Abschluss	
2	1	3	54	*	$\leq 12$	Abbruch	
2	2	1	282	*	13-24	Azubi	
2	2	2	94	*	13-24	Abschluss	
2	2	3	43	*	13-24	Abbruch	
2	3	1	25	*	$\geq 25$	Azubi	
2	3	2	167	*	$\geq 25$	Abschluss	
2	3	3	19	*	$\geq 25$	Abbruch	
2	1	0	92	*	$\leq 12$	Missing	
2	2	0	85	*	13-24	Missing	
2	3	0	81	*	$\geq 25$	Missing	

### 5 NINR-Programm zum Beispiel in Abschnitt 3.3.4, Tabelle 3.7

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 3	* Anzahl manifester Variablen (D, X, Y)
dim 2 2 3 3	* Anzahl Ausprägungen der Variablen R, D, X, Y
lab R D X Y	* Variablenlabel
sub DXY DX	* Subgruppen mit Angaben zu DXY bzw. DX
mod YIDX {XY}	* Logit-Modell zum Ausfalltyp NINR
RIDXY {cov(RD,1), wei(RD), cov(RDY,2), wei(RDY)}	
rec 21	* Anzahl der Datenzeilen in „tab3_5.dat“
rco	* Fallzähler (freq) in „tab3_5.dat“
dat tab3_5.dat	* Datentabelle „tab3_5.dat“, s. o.
des [ 0 0 0 1	* Designmatrix für Effekte: cov(RD,1)
0 0 0 0 0 0 0 0 0 1 0	* cov(RDY,2): R=2, D=2, Y=2
0 0 0 0 0 0 0 0 0 0 1]	* cov(RDY,2): R=2, D=2, Y=3
sta wei(RD) [ 1 1 0 1]	* Startwerte für Effekte RD und RDY:
sta wei(RDY)	* Gewichtung struktureller Nullzellen in den IABS-Daten
[ 1 1 1 1 1 1 0 0 0 1 1 1]	
dum 1 1 1 1	* Referenzkategorien der Variablen R, D, X, Y
wma RDXY tab3_7.fre	* Ausgabe geschätzter Zellenbesetzungen

## 6 MAR- und NINR-Programm zum Beispiel in Abschnitt 3.3.4, Tabelle 3.8

Kommando	Kommentar
res 1	* Anzahl Responsevariablen (R)
man 3	* Anzahl manifester Variablen (D, X, Y)
dim 2 2 3 3	* Anzahl Ausprägungen der Variablen R, D, X, Y
lab R D X Y	* Variablenlabel
sub DXY DX	* Subgruppen mit Angaben zu DXY bzw. DX
mod Y DX {XY}	* Logit-Modell zum Ausfalltyp MAR und NINR
R DXY {cov(RD,1), wei(RD), cov(RDY,2), wei(RDY) cov(RDX,2),wei(RDX)}	
rec 21	* Anzahl der Datenzeilen in „tab3_5.dat“
rco	* Fallzähler (freq) in „tab3_5.dat“
dat tab3_5.dat	* Datentabelle „tab3_5.dat“, s. o.
des [ 0 0 0 1	* Designmatrix für Effekte: cov(RD,1)
0 0 0 0 0 0 0 0 0 1 0	* cov(RDY,2): R=2, D=2, Y=2
0 0 0 0 0 0 0 0 0 0 1	* cov(RDY,2): R=2, D=2, Y=3
0 0 0 0 0 0 0 0 0 1 0	* cov(RDX,2): R=2, D=2, X=2
0 0 0 0 0 0 0 0 0 0 1]	* cov(RDX,2): R=2, D=2, X=3
sta wei(RD) [ 1 1 0 1]	* Startwerte für Effekte RD und RDY:
sta wei(RDY)	* Gewichtung struktureller Nullzellen in den IABS-Daten
[ 1 1 1 1 1 1 0 0 0 1 1 1]	
dum 1 1 1 1	* Referenzkategorien der Variablen R, D, X, Y
wma RDX Y tab3_8.fre	* Ausgabe geschätzter Zellenbesetzungen

## Tabellen

*Tabelle A.1:* Verteilung der unabhängigen Variablen zur Analyse „Auszug aus dem Elternhaus“ in West- und Ostdeutschland (Spaltenprozentwerte)

Variable, Variablengruppe	West- deutschland	Ost-
Gemeindetyp 1996 ( <i>R</i> )		
Kernstadt	23,2	23,5
Ober-/Mittelzentrum	35,8	28,2
Sonstige Gemeinde ( <i>R</i> )	41,0	48,3
Alterskohorte 1996 ( <i>A</i> )		
15 Jahre	12,4	14,0
16 Jahre	12,0	13,0
17 Jahre	11,0	13,9
18 Jahre	11,2	12,4
19 Jahre	9,7	10,6
20 Jahre	8,6	7,4
21 Jahre	7,4	7,3
22 Jahre	6,4	5,9
23 Jahre	6,2	5,5
24 Jahre	6,0	3,8
25 Jahre	5,4	3,6
26 Jahre	3,9	2,7
Geschlecht ( <i>G</i> )		
Männlich ( <i>R</i> )	56,2	58,3
Weiblich	43,8	41,7
Staatsangehörigkeit 1996 ( <i>N</i> )		
Deutsche, EU-Ausländer ( <i>R</i> )	93,8	99,9
nicht EU-Ausländer	6,2	0,1
Haupt-/Nebenwohnsitz 1996 ( <i>H</i> )		
nur Hauptwohnsitz ( <i>R</i> )	94,0	88,5
mit Nebenwohnsitz	6,0	11,5
Schulbesuch 1996 ( <i>B</i> )		
Allgemein bildende Schule ( <i>R</i> )	37,8	37,9
Berufliche Schule	21,6	25,8
Fachhochschule	3,0	2,8
Hochschule	7,5	4,9
kein Schulbesuch	30,0	28,6
Pro-Kopf-Äquivalenzeinkommen 1996 ( <i>E</i> )		
< 948 DM ( <i>R</i> )	7,7	9,6
948-1.359 DM	16,8	25,0
1.359-1.890 DM	26,9	34,6
1.890-2.903 DM	29,6	23,2
2.903+ DM	12,2	3,9
keine Angabe	6,9	3,7
Familientyp 1996 ( <i>F</i> )		
Ehepaar mit ledigen Kindern ( <i>R</i> )	85,6	81,2
Alleinerziehende	14,4	18,8
Geschwisterzahl 1996 ( <i>K</i> )		
keine Geschwister ( <i>R</i> )	34,3	39,1
1 Geschwister	42,8	49,0
2 Geschwister	15,9	9,5
3 und mehr Geschwister	7,0	2,3

Variable, Variablengruppe	West- deutschland	Ost-
Berufliche Stellung der Familienbezugsperson 1996 ( $S^{FB}$ )		
Arbeiter, Angestellte (R)	60,6	71,0
Beamte	10,4	1,3
Selbstständige, mithelfende Familienangehörige	13,2	10,7
Nichterwerbstätige	15,8	17,0
Bildungsniveau der Familienbezugsperson 1996 (CASMIN) ( $C^{FB}$ )		
Hauptschule ohne beruflicher Ausbildung	13,8	1,9
Hauptschule mit beruflicher Ausbildung (R)	44,7	17,4
Mittlere Reife	14,0	54,7
Fachhochschulreife, Abitur	3,7	3,6
Fachhochschulabschluss	6,2	4,7
Hochschulabschluss	10,0	13,8
keine Angabe	7,6	4,0
Quelle des überwiegenden Lebensunterhalts 1996 (L)		
Eltern, Rente, Vermögen, Pflegeversicherung (R)	59,2	47,1
Erwerbstätigkeit, Arbeitslosengeld/-hilfe	39,1	48,4
Sozialhilfe, Sonstige Unterstützung	1,7	4,5
Berufliche Stellung 1996 (S)		
nichterwerbstätig (R)	57,1	53,4
Auszubildende, Wehr- und Zivildienstleistende	16,5	24,0
Arbeiter	12,0	14,0
Angestellte	12,1	6,4
Beamte, Soldaten	1,4	1,7
Selbstständige, mithelfende Familienangehörige	0,9	0,5
Bildungsniveau 1996 (CASMIN) (C)		
Schul-/Hochschulbesuch (R)	39,7	38,9
höchstens Hauptschule ohne beruflicher Ausbildung	12,2	3,4
Hauptschule mit berufl. Ausbildung, Mittlere Reife, Abitur	45,3	55,7
Fachhochschule, Hochschule	1,2	1,1
keine Angabe	1,6	0,9
Zeitveränderliche Variablen	(Anteile bezogen auf Risikopopulationen 1997 bzw. 1998)	
Beendigung Schulbesuch (V)		
Schulbesuch 1996, kein Schulbesuch 1997	(11,0)	(9,0)
Schulbesuch 1996-97, kein Schulbesuch 1998	(12,2)	(11,4)
Schulartwechsel (W)		
Allg. bildende Schule 1996, Berufl. Schule 1997	(7,2)	(9,4)
Allg. bildende Schule 1996-1997, Berufl. Schule 1998	(6,8)	(11,3)
Wechsel Quelle des überwiegenden Lebensunterhalts (U)		
Eltern 1996, Erwerbstätigkeit 1997	(10,8)	(12,7)
Eltern 1996-1997, Erwerbstätigkeit 1998	(11,2)	(12,2)
Betriebswechsel (Z) (Retrospektivfrage)		
Betriebswechsel 1995/1996	6,7	8,9
Betriebswechsel 1996/1997	(9,4)	(10,3)
Betriebswechsel 1997/1998	(11,1)	(13,1)
Insgesamt (Risikopopulation 1996) (n = 100 %)	6.507	1.804
Risikopopulation 1997 insgesamt (n)	(5.685)	(1.572)
Risikopopulation 1998 insgesamt (n)	(4.893)	(1.354)

(R): Referenzkategorie

Quelle: Mikrozensuspanel 1996-1999 (siehe Tab. 4.1); eigene Berechnungen.

Tabelle A.2: Regressionskoeffizienten (b) und Standardfehler (in Klammern) sowie Effektkoeffizienten (exp(b)) des diskreten Hazardraten-Regressionsmodells 21 in Tabelle 4.4

Variable <sup>1)</sup>	Westdeutschland		Ostdeutschland	
	exp(b)	Koeff. <sup>2)</sup>	exp(b)	Koeff. <sup>2)</sup>
Konstanten ( $\alpha_j$ ): Lebensalter (t-15)				
16		-5,7175 * (0,5265)		-4,6143 * (0,4687)
17		-4,8626 * (0,2619)		-4,7320 * (0,3847)
18		-3,9333 * (0,1768)		-4,1011 * (0,2890)
19		-3,3876 * (0,1619)		-3,8766 * (0,2889)
20		-3,1839 * (0,1689)		-3,1863 * (0,2839)
21		-2,9195 * (0,1687)		-3,2146 * (0,3085)
22		-2,8605 * (0,1803)		-2,9838 * (0,3174)
23		-2,7017 * (0,1849)		-3,0450 * (0,3461)
24		-2,3241 * (0,1862)		-2,7722 * (0,3437)
25		-2,3057 * (0,1882)		-2,8074 * (0,3640)
26		-2,1467 * (0,1916)		-2,5902 * (0,3667)
27		-2,0782 * (0,1936)		-2,5811 * (0,3818)
Zeitkonstante Variablen: Status 1996				
Gemeindetyp (R)				
Kernstadt	1,54	0,4349 * (0,0696)	2,06	0,7212 * (0,1238)
Ober-/Mittelzentrum	1,16 1,54	0,1507 * (0,4349) *	1,45 2,06	0,3695 * (0,7212) *
Sonstige Gemeinde (R)	1,00	0,0000	1,00	0,0000
Geschlecht (G)				
männlich (R)	1,00	0,0000	1,00	0,0000
weiblich	1,86	0,6207 * (0,0558)	2,69	0,9886 * (0,1098)
Schulbesuch (B)				
Allgemein bildende Schule (R)	1,00	0,0000	1,00	0,0000
Berufliche Schule	1,66	0,5039 * (0,1200)	2,32	0,8399 * (0,2276)
Fachhochschule	1,86	0,6229 * (0,1563)	2,08	0,7312 * (0,2979)
Hochschule	1,24	0,2141 (0,1338)	2,00	0,6932 * (0,2906)
kein Schulbesuch	1,79	0,5825 * (0,1292)	2,81	1,0333 * (0,2654)



Variable <sup>1)</sup>	Westdeutschland		Ostdeutschland	
	exp(b)	Koeff. <sup>2)</sup>	exp(b)	Koeff. <sup>2)</sup>
Pro-Kopf-Äquivalenzeinkommen ( <i>E</i> )				
< 948 DM ( <i>R</i> )	1,00	0,0000	1,00	0,0000
948-1.359 DM	0,87	-0,1378 (0,1249)	0,67	-0,4018 * (0,1820)
1.359-1.890 DM	0,82	-0,2029 + (0,1185)	0,56	-0,5737 * (0,1811)
1.890-2.903 DM	0,85	-0,1622 (0,1196)	0,59	-0,5336 * (0,1946)
2.903+ DM	1,15	0,1402 (0,1321)	0,74	-0,2995 (0,2908)
keine Angabe	0,83	-0,1875 (0,1556)	0,65	-0,4258 (0,2974)
Familientyp ( <i>F</i> )				
Ehepaar mit ledigen Kindern ( <i>R</i> )	1,00	0,0000	1,00	0,0000
Alleinerziehende	1,24	0,2180 * (0,0758)	1,29	0,2529 * (0,1272)
Geschwisterzahl ( <i>K</i> )				
keine Geschwister ( <i>R</i> )	1,00	0,0000	1,00	0,0000
1 Geschwister	0,97	-0,0339 (0,0621)	1,30	0,2628 * (0,1091)
2 Geschwister	1,01	0,0080 (0,0847)	1,68	0,5214 * (0,1844)
3 und mehr Geschwister	1,40	0,3339 * (0,1127)	3,27	1,1850 * (0,2669)
Berufliche Stellung Familienbezugsperson ( <i>S<sup>FB</sup></i> )				
Arbeiter, Angestellte ( <i>R</i> )	1,00	0,0000	1,00	0,0000
Beamte	0,95	-0,0510 (0,0924)	1,17	0,1557 (0,3838)
Selbstständige, mithelfende Familienangehörige	0,77	-0,2599 * (0,0897)	1,20	0,1804 (0,1618)
Nichterwerbstätige	0,87	-0,1426 + (0,0788)	1,06	0,0590 (0,1358)
Berufliche Stellung ( <i>S</i> )				
nichterwerbstätig ( <i>R</i> )	1,00	0,0000	1,00	0,0000
Auszubildende, Wehr- und Zivildienstleistende	1,05	0,0456 (0,1074)	1,05	0,0465 (0,1888)
Arbeiter	0,98	-0,0202 (0,1071)	0,80	-0,2225 (0,2031)
Angestellte	1,27	0,2384 * (0,0993)	0,92	-0,0840 (0,2263)
Beamte, Soldaten	1,52	0,4214 * (0,1993)	2,56	0,9401 * (0,2981)
Selbstständige, mithelfende Familienangehörige	0,87	-0,1422 (0,3400)	1,78	0,5742 (0,6554)
Zeitveränderliche Variablen <sup>3)</sup>				
Referenzkategorien: jeweils „ohne Wechsel“				
Betriebswechsel ( <i>Z</i> )				
Betriebswechsel 1995/1996	1,06	0,0584 (0,0961)	1,46	0,3756 * (0,1560)

Variable <sup>1)</sup>	Westdeutschland		Ostdeutschland	
	exp(b)	Koeff. <sup>2)</sup>	exp(b)	Koeff. <sup>2)</sup>
Betriebswechsel 1996/1997	1,25	0,2200 * (0,0982)	1,08	0,0758 (0,1756)
Betriebswechsel 1997/1998	1,37	0,3150 * (0,1311)	1,05	0,0486 (0,2436)
Schulbesuch/kein Schulbesuch (V)				
Schulbesuch 1996, kein Schulbesuch 1997	1,28	0,2441 * (0,1069)	1,50	0,4076 + (0,2145)
Schulbesuch 1996-97, kein Schulbesuch 1998	1,47	0,3830 * (0,1323)	1,62	0,4839 + (0,2474)
Schulartwechsel (W)				
Allgemein bildende Schule 1996, Berufliche Schule 1997	1,47	0,3829 * (0,1607)	1,42	0,3540 (0,2866)
Allgemein bildende Schule 1996-97, Berufliche Schule 1998	1,67	0,5111 * (0,2285)	0,98	-0,0233 (0,4346)

Für weitere Angaben siehe Tabelle 4.4.

- 1) Jeweilige Referenzkategorien sind mit (R) bezeichnet.
- 2) Bei statistisch signifikant gekennzeichneten Koeffizienten steht \* für  $b/s.e. > |1,96|$  bzw. + für  $b/s.e. > |1,65|$ .
- 3) Referenzkategorien wie zuvor und „ohne Wechsel“.

**Tabelle A.3:** Regressionskoeffizienten und Standardfehler der log-linearen Modelle in Tabelle 5.8 zum Ausfall und Status von 1997 bis 1998 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 (Logit-Spezifikation)

Variablen	Modell			
	1 (MCAR)	8 (MAR)	12 (NINR)	13 (NINR)
Strukturmodell				
Übergang 1996/97				
$Y_{97}=ABG, Y_{96}=GYO$	-1,7025 * (0,2241)	-1,7025 * (0,2241)	-1,6267 * (0,2251)	-1,6161 * (0,2271)
$Y_{97}=ABI, Y_{96}=GYO$	-3,8180 * (0,4705)	-3,8180 * (0,4705)	-3,8040 * (0,4708)	-3,8075 * (0,4714)
$Y_{97}=ABG, A=17$	-0,3501 (0,2450)	-0,3501 (0,2450)	-0,3591 (0,2439)	-0,3501 (0,2444)
$Y_{97}=ABG, A=18-19$	-0,4691 (0,2504)	-0,4691 (0,2504)	-0,4302 (0,2500)	-0,4374 * (0,2530)
$Y_{97}=ABI, A=17$	0,8459 (0,5078)	0,8459 (0,5078)	0,8416 (0,5078)	0,8473 (0,5080)
$Y_{97}=ABI, A=18-19$	3,3439 * (0,4662)	3,3439 * (0,4662)	3,3481 * (0,4663)	3,3381 * (0,4667)
$Y_{97}=ABG, L=Ost$	-0,2029 (0,2765)	-0,2029 (0,2765)	-0,2050 (0,2747)	-0,2139 (0,2752)
$Y_{97}=ABI, L=Ost$	0,9658 * (0,2333)	0,9658 * (0,2333)	0,9732 * (0,2333)	0,9716 * (0,2336)
$Y_{97}=ABG, S=NET$	0,4096 (0,3334)	0,4096 (0,3334)	0,4478 (0,3305)	0,4380 (0,3305)
$Y_{97}=ABG, S=ARB$	0,8993 * (0,2351)	0,8993 * (0,2351)	0,9412 * (0,2340)	0,9240 * (0,2341)
$Y_{97}=ABG, S=SEL$	-0,3969 (0,3411)	-0,3969 (0,3411)	-0,4595 (0,3410)	-0,4695 (0,3414)
$Y_{97}=ABG, S=BEA$	-1,1233 * (0,4471)	-1,1233 * (0,4471)	-1,1322 * (0,4457)	-1,1311 * (0,4453)
$Y_{97}=ABI, S=NET$	-0,0211 (0,2773)	-0,0211 (0,2773)	-0,0153 (0,2778)	-0,0179 (0,2776)
$Y_{97}=ABI, S=ARB$	-0,0057 (0,2282)	-0,0057 (0,2282)	-0,0029 (0,2288)	(-0,0315)
$Y_{97}=ABI, S=SEL$	-0,1379 (0,2461)	-0,1379 (0,2461)	-0,1468 (0,2460)	-0,1443 (0,2462)
$Y_{97}=ABI, S=BEA$	-0,1124 (0,2421)	-0,1124 (0,2421)	-0,1147 (0,2420)	-0,1132 (0,2422)
Übergang 1997/98				
$Y_{98}=ABG, Y_{97}=GYO$	-2,4561 * (0,3497)	-2,4561 * (0,3497)	-2,4169 * (0,3505)	-2,2360 * (0,3581)
$Y_{98}=ABI, Y_{97}=GYO$	-3,3004 * (0,3621)	-3,3004 * (0,3621)	-3,2936 * (0,3628)	-3,2632 * (0,3679)
$Y_{98}=ABG, A=17$	-0,7127 (0,3871)	-0,7127 (0,3871)	-0,7177 (0,3868)	-0,6730 (0,3811)
$Y_{98}=ABG, A=18-19$	0,4720 (0,4156)	0,4720 (0,4156)	0,4852 (0,4157)	0,5646 (0,4171)

Variablen	Modell			
	1 (MCAR)	8 (MAR)	12 (NINR)	13 (NINR)
$Y_{98}=ABI, A=17$	1,6175 * (0,3464)	1,6175 * (0,3464)	1,6179 * (0,3464)	1,6404 * (0,3470)
$Y_{98}=ABI, A=18-19$	4,3331 * (0,3711)	4,3331 * (0,3711)	4,3375 * (0,3713)	4,3767 * (0,3724)
$Y_{98}=ABG, L=Ost$	0,5704 (0,4548)	0,5704 (0,4548)	0,5764 (0,4542)	0,5872 (0,4450)
$Y_{98}=ABI, L=Ost$	2,2227 * (0,2940)	2,2227 * (0,2940)	2,2250 * (0,2940)	2,2446 * (0,2935)
$Y_{98}=ABG, S=NET$	0,8153 (0,5308)	0,8153 (0,5308)	0,8502 (0,5317)	0,9946 (0,5227)
$Y_{98}=ABG, S=ARB$	1,0832 * (0,3957)	1,0832 * (0,3957)	1,0986 * (0,3955)	1,1413 * (0,3873)
$Y_{98}=ABG, S=SEL$	-0,2385 (0,5877)	-0,2385 (0,5877)	-0,2612 (0,5878)	-0,3789 (0,5848)
$Y_{98}=ABG, S=BEA$	-0,5081 (0,6554)	-0,5081 (0,6554)	-0,5054 (0,6554)	-0,5073 (0,6503)
$Y_{98}=ABI, S=NET$	-0,1768 (0,3561)	-0,1768 (0,3561)	-0,1694 (0,3567)	-0,1305 (0,3590)
$Y_{98}=ABI, S=ARB$	-0,2310 (0,2784)	-0,2310 (0,2784)	-0,2296 (0,2784)	-0,2340 (0,2778)
$Y_{98}=ABI, S=SEL$	0,5047 (0,2758)	0,5047 (0,2758)	0,4990 (0,2764)	0,4596 (0,2801)
$Y_{98}=ABI, S=BEA$	0,4085 (0,2854)	0,4085 (0,2854)	0,4089 (0,2852)	0,4096 (0,2835)
Übergang 1998/99				
$Y_{99}=ABG, Y_{98}=GYO$	-1,8450 * (0,4071)	-1,8450 * (0,4071)	-1,8449 * (0,4070)	-1,8550 * (0,4224)
$Y_{99}=ABI, Y_{98}=GYO$	-0,3615 (0,2478)	-0,3615 (0,2478)	-0,3615 (0,2478)	-0,4243 (0,2692)
$Y_{99}=ABG, A=17$	1,3090 * (0,4483)	1,3090 * (0,4483)	1,3090 * (0,4483)	1,3071 * (0,4510)
$Y_{99}=ABG, A=18-19$	2,2664 * (0,6023)	2,2664 * (0,6023)	2,2663 * (0,6022)	2,2625 * (0,6034)
$Y_{99}=ABI, A=17$	1,8844 * (0,2811)	1,8844 * (0,2811)	1,8844 * (0,2810)	1,8587 * (0,2862)
$Y_{99}=ABI, A=18-19$	2,2928 * (0,4585)	2,2928 * (0,4585)	2,2928 * (0,4585)	2,2706 * (0,4607)
$Y_{99}=ABG, L=Ost$	0,6484 (0,7426)	0,6484 (0,7426)	0,6484 (0,7426)	0,6522 (0,7445)
$Y_{99}=ABI, L=Ost$	1,1687 * (0,4736)	1,1687 * (0,4736)	1,1687 * (0,4736)	1,1541 * (0,4758)
$Y_{99}=ABG, S=NET$	0,0754 (0,7046)	0,0754 (0,7046)	0,0754 (0,7046)	0,0648 (0,7046)
$Y_{99}=ABG, S=ARB$	-0,3042 (0,5771)	-0,3042 (0,5771)	-0,3042 (0,5771)	-0,3049 (0,5809)
$Y_{99}=ABG, S=SEL$	-0,1204 (0,5486)	-0,1204 (0,5486)	-0,1204 (0,5486)	-0,1188 (0,5517)

Variablen	Modell			
	1 (MCAR)	8 (MAR)	12 (NINR)	13 (NINR)
Y <sub>99</sub> =ABG, S=BEA	-0,0379 (0,5725)	-0,0379 (0,5725)	-0,0379 (0,5725)	-0,0290 (0,5771)
Y <sub>99</sub> =ABI, S=NET	-0,1476 (0,4987)	-0,1476 (0,4987)	-0,1476 (0,4987)	-0,1510 (0,4967)
Y <sub>99</sub> =ABI, S=ARB	-0,5273 (0,3839)	-0,5273 (0,3839)	-0,5273 (0,3839)	-0,5578 (0,3870)
Y <sub>99</sub> =ABI, S=SEL	-0,3937 (0,3716)	-0,3937 (0,3716)	-0,3937 (0,3716)	-0,3571 (0,3763)
Y <sub>99</sub> =ABI, S=BEA	0,1636 (0,3644)	0,1636 (0,3644)	0,1636 (0,3644)	0,2063 (0,3717)
Ausfallmodell				
Übergang 1996/97 R <sub>97</sub> =AUS	-0,1807 (0,1348)	0,1496 (0,4125)	0,1258 (0,4118)	0,1200 (0,4130)
R <sub>97</sub> =AUS, A=17		-0,6542 (0,4617)	-0,6439 (0,4624)	-0,6548 (0,4634)
R <sub>97</sub> =AUS, A=18-19		-0,4181 (0,4106)	-0,3946 (0,4110)	-0,3878 (0,4129)
R <sub>97</sub> =AUS, L=Ost		-0,4328 (0,3530)	-0,4120 (0,3526)	-0,4034 (0,3538)
R <sub>97</sub> =AUS, S=NET		-0,0542 (0,4174)	-0,0418 (0,4178)	-0,0377 (0,4179)
R <sub>97</sub> =AUS, S=ARB		1,0795 * (0,3672)	1,0756 * (0,3672)	1,0807 * (0,3675)
R <sub>97</sub> =AUS, S=SEL		-0,3410 (0,5890)	-0,3331 (0,5892)	-0,3258 (0,5894)
R <sub>97</sub> =AUS, S=BEA		-0,3501 (0,4327)	-0,3414 (0,4326)	-0,3395 (0,4323)
Übergang 1997/98 R <sub>98</sub> =AUS	-0,1809 (0,1231)	-0,3193 (0,3954)	-0,5092 (0,4066)	-0,8663 (0,5064)
R <sub>98</sub> =AUS, A=17		-0,0364 (0,4254)	0,0009 (0,4277)	0,0062 (0,4449)
R <sub>98</sub> =AUS, A=18-19		0,2945 (0,3979)	0,3148 (0,4144)	0,1652 (0,5236)
R <sub>98</sub> =AUS, L=Ost		0,2686 (0,3147)	0,2736 (0,3195)	0,2095 (0,3399)
R <sub>98</sub> =AUS, S=NET		0,4505 (0,3982)	0,4103 (0,4003)	0,3556 (0,4049)
R <sub>98</sub> =AUS, S=ARB		-0,7013 * (0,3488)	-0,8007 * (0,3540)	-0,8909 * (0,3623)
R <sub>98</sub> =AUS, S=SEL		-0,5313 (0,4353)	-0,5128 (0,4366)	-0,5140 (0,4381)
R <sub>98</sub> =AUS, S=BEA		0,4835 (0,3913)	0,5488 (0,3930)	0,5665 (0,3958)
R <sub>98</sub> =AUS, Y <sub>97</sub> =ABG			0,8049 * (0,3418)	

Variablen	Modell			
	1 (MCAR)	8 (MAR)	12 (NINR)	13 (NINR)
$R_{98}=AUS, Y_{97}=ABI$			0,2423 (0,2593)	
$R_{98}=AUS, Y_{98}=ABG$				1,3819 * (0,5967)
$R_{98}=AUS, Y_{98}=ABI$				0,6745 (0,6346)
Übergang 1998/99				
$R_{99}=AUS$	-1,8713 * (0,0892)	-2,4306 * (0,2685)	-2,5339 * (0,2820)	-2,2877 * (0,4617)
$R_{99}=AUS, A=17$		0,4849 (0,2868)	0,5119 (0,2909)	0,6219 (0,3448)
$R_{99}=AUS, A=18-19$		0,6415 * (0,2722)	0,6519 * (0,3141)	0,8491 * (0,3935)
$R_{99}=AUS, L=Ost$		0,4637 * (0,2333)	0,4640 (0,2414)	0,5262 * (0,2467)
$R_{99}=AUS, S=NET$		0,0857 (0,3231)	0,0644 (0,3243)	0,0575 (0,3246)
$R_{99}=AUS, S=ARB$		0,2696 (0,2306)	0,2158 (0,2347)	0,2097 (0,2363)
$R_{99}=AUS, S=SEL$		-0,1995 (0,2824)	-0,1913 (0,2828)	-0,1974 (0,2831)
$R_{99}=AUS, S=BEA$		-0,3496 (0,3077)	-0,3264 (0,3085)	-0,3211 (0,3089)
$R_{99}=AUS, Y_{98}=ABG$			0,3866 (0,2828)	
$R_{99}=AUS, Y_{98}=ABI$			0,0834 (0,2378)	
$R_{99}=AUS, Y_{99}=ABG$				-0,0522 (0,6248)
$R_{99}=AUS, Y_{99}=ABI$				-0,3990 (0,6761)
Log-Likelihood	-5.426,92	-5.391,75	-5.384,22	-5.383,96
Parameter Gesamtmodell	80	101	105	105
Pseudo-R <sup>2</sup> Strukturmodell				
Übergang 1996/97	0,1467	0,1467	0,1471	0,1453
Übergang 1997/98	0,5598	0,5598	0,5672	0,5538
Übergang 1998/99	0,6235	0,6235	0,6343	0,6451
Pseudo-R <sup>2</sup> Ausfallmodell		0,0299	0,0411	0,0512
Stichprobengröße	1.309			
Ausfälle (kumuliert) in %				
Übergang 1996/97	7,7			
Übergang 1997/98	17,0			
Übergang 1998/99	28,0			

\* Signifikant mit  $p < 0,05$

Variablen (0 = Referenzkategorie)

Y <sub>i</sub>	Status	
	GYO	Besuch der gymnasialen Oberstufe (0)
	ABG	Abgang (Berufliche Schule, kein Schulbesuch)
	ABI	Abschluss (Allg. und Fachgeb. Hochschulreife inkl. Fachhochschulreife, FH-/Hochschulbesuch)
R <sub>i</sub>	Responseindikator	
	kein Ausfall (0)	
	AUS	Ausfall
A	Alterskohorte zum Zeitpunkt 1996	
	16 Jahre (0)	
	17 Jahre	
	18-19 Jahre	
L	Bundesland/Regelschulzeit	
	WEST	13 Jahre: Westdeutschland inkl. Berlin und Brandenburg (0)
	OST	12 Jahre: Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen
S	Stellung im Beruf des Familienvorstands	
	NET	Nichterwerbstätige
	ARB	Arbeiter
	SEL	Selbstständige inkl. Mithelfende Familienangehörige
	BEA	Beamte
	ANG	Angestellte (0)



Tabelle A.4: Status der 1996 15- bis 24-jährigen Auszubildenden im April 1997 in der Beschäftigtenstichprobe und im Mikrozensuspanel – in Prozent

Merkmal	Insgesamt (n)	ohne Ausfall (%)	darunter (ohne Ausfall = 100 %)			mit Ausfall (%)
			Abschluss	Abbruch	Auszubi.	
<b>Beschäftigtenstichprobe</b>						
<b>Mikrozensuspanel</b>						
Insgesamt B	23.044	100,0	22,1	13,6	64,3	
M	1.429	81,0	25,5	10,6	63,9	19,0
Betriebsort (Land)						
West B	17.667	100,0	23,2	13,4	63,4	
M	1.054	82,1	27,1	11,6	61,4	17,9
Ost (+Berlin) B	5.377	100,0	18,3	14,4	67,3	
M	375	78,1	20,8	7,9	71,3	21,9
Staatsangeh. (West)						
Deutsche bzw. Insg. B	21.252	100,0	22,4	13,2	64,5	
M	1.310	80,8	25,9	9,8	64,3	19,2
Ausländer B	1.818	100,0	18,8	19,1	62,0	
M	119	83,2	21,2	19,2	59,6	16,8
Alter 1996						
15-18 B	7.486	100,0	5,9	7,2	87,0	
M	536	88,1	8,1	9,8	82,2	11,9
19-20 B	9.838	100,0	28,2	15,3	56,5	
M	591	78,7	33,6	10,1	56,3	21,3
21-22 B	4.296	100,0	31,8	18,9	49,2	
M	235	76,2	43,6	14,0	42,5	23,8
23-24 B	1.424	100,0	35,4	19,9	44,7	
M	67	62,7	54,8	11,9	33,3	37,3
Geschlecht						
männlich B	13.304	100,0	19,9	14,8	65,2	
M	834	85,7	23,5	10,1	66,4	14,3
weiblich B	9.740	100,0	25,0	11,9	63,0	
M	595	74,5	28,7	11,5	59,8	25,6
Allgem. Abschluss						
<Abitur B	21.083	100,0	21,6	13,9	64,5	
M	1.251	81,9	25,5	10,2	64,4	18,2
Abitur B	1.961	100,0	27,4	10,4	62,2	
M	178	75,3	25,4	14,2	60,5	24,7
Beruf 1996						
Gruppe 1 B	8.063	100,0	19,6	14,9	65,5	
M	482	86,9	25,5	9,6	64,9	13,1
Gruppe 2 B	1.608	100,0	17,7	15,9	66,5	
M	113	83,2	16,0	11,7	72,3	16,8
Gruppe 3 B	2.368	100,0	19,0	15,9	65,2	
M	141	85,8	19,8	13,2	66,9	14,2
Gruppe 4 B	3.871	100,0	26,7	11,4	61,9	
M	248	74,2	31,5	11,4	57,1	25,8
Gruppe 5 B	5.467	100,0	25,5	10,3	64,2	
M	341	80,1	26,7	11,4	61,9	19,9
Gruppe 6 B	1.667	100,0	20,6	18,1	61,4	
M	104	64,4	26,9	6,0	67,2	35,6
Besch.dauer 1996						
1-12 Monate B	9.944	100,0	4,0	8,4	87,6	
M	585	84,3	3,9	11,0	85,2	15,7

Merkmal	Insgesamt (n)	ohne Ausfall (%)	darunter (ohne Ausfall = 100 %)			mit Ausfall (%)
			Abschluss	Abbruch	Auszubi.	
<b>Beschäftigtenstichprobe</b>						
<b>Mikrozensuspanel</b>						
13-24 Monate B	7.697	100,0	17,8	11,3	70,8	
M	504	83,1	22,4	10,3	67,3	16,9
25+ Monate B	5.403	100,0	61,4	26,4	12,2	
M	292	72,3	79,2	9,0	11,9	27,7
ohne Ang. M	48	72,9	42,9	20,0	37,1	27,1

Tabelle zu Abbildung 6.2 im Textteil.

Tabelle A.5: Regressionskoeffizienten und Standardfehler ausgewählter log-linearer Modelle zum Status und Ausfall 1996/97 der bis 24-jährigen Auszubildenden des Jahres 1996

Teilmodell (Logit-Spezifikation)	Modell			
	2 (MCAR)	18 (MAR)	19 (NINR)	20 (NINR)
<b>Strukturmodell</b>				
Konstante Y=2 Abschluss	-4,0896 *	-4,0896 *	-4,0895 *	-4,0889 *
	(0,0755)	(0,0755)	(0,0755)	(0,0755)
Alter (1996)=19-20 Jahre	0,8925 *	0,8925 *	0,8935 *	0,8932 *
	(0,0618)	(0,0618)	(0,0618)	(0,0618)
Alter (1996)=21-22 Jahre	1,1615 *	1,1616 *	1,1656 *	1,1658 *
	(0,0722)	(0,0722)	(0,0722)	(0,0722)
Alter (1996)=23-24 Jahre	1,2371 *	1,2371 *	1,2390 *	1,2398 *
	(0,0952)	(0,0952)	(0,0951)	(0,0951)
Geschlecht=weiblich	0,0898	0,0898	0,0856	0,0860
	(0,0531)	(0,0531)	(0,0531)	(0,0531)
Land/Betriebsstätte=Ost	-0,4663 *	-0,4663 *	-0,4667 *	-0,4679 *
	(0,0529)	(0,0529)	(0,0528)	(0,0528)
Allg. Bildungsabschluss=Abitur	0,3476 *	0,3476 *	0,3358 *	0,3352 *
	(0,0731)	(0,0731)	(0,0730)	(0,0731)
Staatsangeh.(West)=Ausländer	-0,3948 *	-0,3948 *	-0,3921 *	-0,3922 *
	(0,0811)	(0,0811)	(0,0810)	(0,0810)
Job 1996=Bauberufe	0,4264 *	0,4264 *	0,4194 *	0,4204 *
	(0,0920)	(0,0920)	(0,0918)	(0,0918)
Job=Raumausstatter-Techniker	0,3326 *	0,3326 *	0,3305 *	0,3306 *
	(0,0769)	(0,0769)	(0,0768)	(0,0768)
Job=Waren-/Dienstleist.kaufleute	0,8372 *	0,8372 *	0,8282 *	0,8285 *
	(0,0681)	(0,0681)	(0,0680)	(0,0680)
Job=Org.-/Erziehungsberufe	0,3743 *	0,3744 *	0,3705 *	0,3703 *
	(0,0671)	(0,0671)	(0,0670)	(0,0670)
Job=Verk./Allg. Dienstleist.ber.	0,6284 *	0,6284 *	0,6142 *	0,6137 *
	(0,0935)	(0,0935)	(0,0934)	(0,0934)
T Beschäft.dauer 1996=13-24 M	1,5994 *	1,5994 *	1,5954 *	1,5965 *
	(0,0599)	(0,0599)	(0,0598)	(0,0598)
T Beschäft.dauer 1996=25+ M	4,6274 *	4,6274 *	4,6118 *	4,6185 *
	(0,0701)	(0,0701)	(0,0702)	(0,0701)
<b>Konstante Y=3 Abbruch</b>				
Konstante Y=3 Abbruch	-2,8365 *	-2,8365 *	-2,8351 *	-2,8431 *
	(0,0588)	(0,0588)	(0,0588)	(0,0589)
Alter (1996)=19-20 Jahre	0,5747 *	0,5747 *	0,5743 *	0,5716 *
	(0,0569)	(0,0569)	(0,0569)	(0,0569)
Alter (1996)=21-22 Jahre	1,1072 *	1,1072 *	1,1059 *	1,1038 *
	(0,0680)	(0,0680)	(0,0680)	(0,0680)
Alter (1996)=23-24 Jahre	1,2910 *	1,2910 *	1,2873 *	1,2849 *
	(0,0933)	(0,0933)	(0,0951)	(0,0933)
Geschlecht=weiblich	-0,0845	-0,0845	-0,0851	-0,0865
	(0,0561)	(0,0561)	(0,0562)	(0,0561)
Land/Betriebsstätte=Ost	-0,0316	-0,0316	-0,0326	-0,0370
	(0,0520)	(0,0520)	(0,0521)	(0,0520)
Allg. Bildungsabschluss=Abitur	-0,2858 *	-0,2858 *	-0,2792 *	-0,2876 *
	(0,0859)	(0,0859)	(0,0860)	(0,0860)
Staatsangeh.(West)=Ausländer	0,2684 *	0,2684 *	0,2663 *	0,2634 *
	(0,0734)	(0,0734)	(0,0734)	(0,0734)

Teilmodell (Logit-Spezifikation)	Modell			
	2 (MCAR)	18 (MAR)	19 (NINR)	20 (NINR)
Job 1996=Bauberufe	0,3761 * (0,0848)	0,3761 * (0,0848)	0,3776 * (0,0848)	0,3779 * (0,0848)
Job=Raumausstatter-Techniker	0,3249 * (0,0723)	0,3249 * (0,0723)	0,3263 * (0,0722)	0,3273 * (0,0722)
Job=Waren-/Dienstleist.kaufleute	0,1574 * (0,0723)	0,1574 * (0,0723)	0,1622 * (0,0723)	0,1657 * (0,0722)
Job=Org.-/Erziehungsberufe	-0,1832 * (0,0714)	-0,1832 * (0,0714)	-0,1804 * (0,0714)	-0,1781 * (0,0713)
Job=Verk./Allg. Dienstleist.ber.	0,6192 * (0,0883)	0,6193 * (0,0883)	0,6243 * (0,0883)	0,6251 * (0,0882)
T Beschäft.dauer 1996=1-2 Jahre	0,3483 * (0,0517)	0,3483 * (0,0517)	0,3494 * (0,0517)	0,3574 * (0,0524)
T Beschäft.dauer 1996=2+ Jahre	2,8242 * (0,0626)	2,8242 * (0,0626)	2,8269 * (0,0627)	2,8513 * (0,0632)
Ausfallmodell (nur für Mikrozensuspanel D=2)				
Konstante R=0 Ausfall	-1,4708 * (0,0690)	-2,6703 * (0,2056)	-2,7794 * (0,2184)	-2,6532 * (0,2210)
A=19-20 Jahre		0,5996 * (0,1827)	0,6689 * (0,1873)	0,6896 * (0,1901)
A=21-22 Jahre		0,7150 * (0,2219)	0,8839 * (0,2357)	0,8751 * (0,2424)
A=23-24 Jahre		1,4352 * (0,3225)	1,7451 * (0,3628)	1,7503 * (0,3671)
G=weiblich		0,5809 * (0,1785)	0,6178 * (0,1906)	0,6532 * (0,1962)
L=Ost		0,3916 * (0,1605)	0,4108 * (0,1714)	0,3686 * (0,1753)
J=Bauberufe		0,3488 (0,3030)	0,1916 (0,3191)	0,2916 (0,3254)
J=Raumausstatter-Techniker		-0,0574 (0,2930)	-0,2230 (0,3190)	-0,2413 (0,3298)
J=Waren-/Dienstleist.kaufleute		0,4857 * (0,2259)	0,5707 * (0,2419)	0,5700 * (0,2467)
J=Org.-/Erziehungsberufe		0,0765 (0,2324)	0,1039 (0,2453)	0,1225 (0,2509)
J=Verk.-/Allg. Dienstleist.berufe		0,9978 * (0,2781)	1,0629 * (0,2972)	1,0387 * (0,3055)
T=13-24 Monate		-0,0020 (0,1720)	0,2021 (0,1769)	0,0658 (0,2152)
T=25+ Monate		0,4412 * (0,1902)	1,5104 * (0,2964)	0,5912 (0,4669)
Y=Abschluss (2)			-2,1535 * (0,5118)	-1,5165 * (0,5713)
Y=Abbruch (3)			0,4042 (0,3251)	-1,5471 (2,0141)
Y=Abbruch (3) * T=13-24 M				1,6181 (2,1396)
Y=Abbruch (3) * T=25+ Monate				3,2299 (2,0955)

Teilmodell (Logit-Spezifikation)	Modell			
	2 (MCAR)	18 (MAR)	19 (NINR)	20 (NINR)
Log-Likelihood	-145.865,3	-145.819,4	-145.802,9	-145.798,3
	2	4	4	7
Parameter	31	43	45	47

Tabelle zu Tabelle 6.4 im Textteil.

\* Signifikant mit  $p < 0,05$

Stichprobengröße (n): D=1 Beschäftigtenstichprobe (23.044)

D=2 Mikrozensuspanel (1.381, darunter 258 Ausfälle)

Variablen und Referenzkategorien:

Y Status 1997: Auszubildender (1)

R Responseindikator: kein Ausfall (1) [strukturelle Nullzelle für D=1 & R=0]

D Datentyp: Beschäftigtenstichprobe (1)

A Altersgruppe zum Zeitpunkt 1996: 15-18 Jahre (1)

G Geschlecht: männlich (1)

L Bundesland der Betriebsstätte: Westdeutschland (ohne Berlin) (1)

B Allg. Bildungsabschluss: nicht Abitur (inkl. ohne Abschluss, ohne Angaben) (1)

S Staatsangehörigkeit (nur Westdeutschland): Deutsche (1)

J Ausbildungsberuf 1996: Pflanzenbauer bis Ernährungsberufe (1)

T Beschäftigungsdauer beim gegenw. Arbeitgeber / Betriebsstätte (1996): =12 Monate (1)

Tabelle A.6: Regressionskoeffizienten gewichteter Logit-Modelle zum Statusübergang 1996/97 der bis 24-jährigen Auszubildenden des Jahres 1996

Koeffizient	Beschäftig- tenstichprobe ungew.	Mikro- zensuspanel ungew.	Querschn. faktor		Querschn. + Logit		Längsschn. faktor		Längsschn. + Mobilität	
			PW	Gewicht=g1	PW	Gewicht=g2	PW	Gewicht=g3	PW	Gewicht=g4
Konstante Y=2 Abschluss	-4,1094 *	-3,8894 *		-3,8971 *		-3,8080 *		-3,8034 *		-3,8312 *
Alter (1996)=19-20 Jahre	0,9052 *	0,8883 *		0,8629 *		0,8555 *		0,9288 *		0,9241 *
Alter (1996)=21-22 Jahre	1,1598 *	1,1709 *		1,1715 *		1,7310 *		1,7845 *		1,7271 *
Alter (1996)=23-24 Jahre	1,2382 *	2,2786 *		2,2171 *		2,0708 *		2,2205 *		1,9724 *
Geschlecht=weiblich	0,0954	-0,0091		-0,0030		-0,0159		-0,1297		-0,1423
Land/Betriebsstätte=Ost	-0,4704 *	-0,4050		-0,4053		-0,3742		-0,3519		-0,3714
Allgem. Bildungsabschluss=Abitur	0,3640 *	-0,4415		-0,4077		-0,4092		-0,4263	**	-0,3176
Staatsangeh.(West)=Ausländer	-0,3967 *	-0,7000		-0,5722		-0,6670		-0,7177 *		-0,6448
Job/Berufsgruppe 1996=Bauberufe	0,4808 *	-0,6922		-0,5494		-0,6381		-0,7213	**	-0,7319 **
Job=Braunmausstatter-Techniker	0,3577 *	-0,1868		-0,2167		-0,2941		-0,2012		-0,2722
toJob=Waren-/Dienstleist.kaufleute	0,8398 *	0,9185 *		0,8398 *		0,7371 *		0,8140 *		0,7781 *
Job=Organisations-/Erziehungsberufe	0,3709 *	0,5013		0,4430		0,4283		0,3939		0,4032
Job=Verk.-/Allg. Dienstleist.berufe	0,6457 *	0,4003		0,4482		0,4383		0,5733		0,6064
T Beschäft.dauer 1996=13-24 M.	1,5764 *	1,9163 *		1,9641 *		1,9230 *		1,8949 *		1,9301 *
T Beschäft.dauer 1996=25+ Monate	4,6191 *	4,8352 *		4,8249 *		4,7313 *		4,6597 *		4,6839 *
Konstante Y=3 Abbruch	-2,8712 *	-2,3419 *		-2,3040 *		-2,2836 *		-2,3490 *	**	-2,3402 *
Alter (1996)=19-20 Jahre	0,5986 *	0,0992		0,0670		0,0360		0,0706	**	0,0755 *
Alter (1996)=21-22 Jahre	1,1316 *	0,5874		0,4963		0,5348		0,5664		0,5149
Alter (1996)=23-24 Jahre	1,3225 *	0,4273		0,5306		0,3775		0,5781		0,3389
Geschlecht=weiblich	-0,0970	0,2124		0,2396		0,2677		0,1984		0,1327
Land/Betriebsstätte=Ost	-0,0216	-0,3677		-0,3757		-0,3548		-0,2869		-0,2780
Allgem. Bildungsabschluss=Abitur	-0,3162 *	0,2444		0,2572		0,2210		0,1939		0,2035
Staatsangeh.(West)=Ausländer	0,2396 *	0,7273 *		0,8158 *		0,7892 *		0,7108 *		0,7293 *
Job/Berufsgruppe 1996=Bauberufe	0,3973 *	0,2535		0,2326		0,2166		0,3104		0,2656
Job=Braunmausstatter-Techniker	0,3383 *	0,2320		0,2330		0,3209		0,3896		0,3048
Job=Waren-/Dienstleist.kaufleute	0,1650 *	0,1033		-0,0323		-0,0826		0,0068		0,0781
Job=Organisations-/Erziehungsberufe	-0,1946 *	0,0014		-0,0249		-0,0599		-0,0276		0,0064

Tabelle A.6: Regressionskoeffizienten gewichteter Logit-Modelle zum Statusübergang 1996/97 der bis 24-jährigen Auszubildenden des Jahres 1996

Koeffizient	Beschäftig- tenstichprobe ungew.	Mikro- zensuspanel ungew.	Querschn. faktor		Querschn. + Logit		Längsschn. faktor		Längsschn. + Mobilität	
			PW	Gewicht=g1	PW	Gewicht=g2	PW	Gewicht=g3	PW	Gewicht=g4
Job=Verk./Allg. Dienstleist. berufe	0,6622 *	-0,5610	**	-0,4635		-0,6307	**	-0,3972		-0,4112
T Beschäft.dauer 1996=13-24 M.	0,3600 *	0,1839		0,2266		0,2184		0,2254		0,2170
T Beschäft.dauer 1996=25+ Monate	2,8601 *	1,7976 *	**	1,8120 *	**	1,7451 *	**	1,6892 *	**	1,6694 *
Wald-Test (df=30)		87,77		84,46		90,38		78,78		77,44
P		0,00		0,00		0,00		0,00		0,00

Tabelle zu Tabelle 6.5 im Textteil.

\*) Koeffizient signifikant mit  $P < 0,05$ PW = \*\*: Differenz zwischen Beschäftigtenstichprobe und Mikrozensuspanel signifikant mit  $p < 0,05$  (nicht für Mehrfachtests korrigierte Werte)



*Tabelle A.7: Berufswechsel von Absolventen einer Berufsausbildung im dualen System der Übergänge 1996/97, 1997/98 und 1998/99 im Mikrozensuspanel und in der Beschäftigtenstichprobe nach verschiedenen Merkmalen*

Merkmal (ein Jahr vor dem Abschluss)	Mikrozensuspanel			Beschäftigtenstichprobe				
	Insgesamt		darunter mit Berufswechsel 95 %-Konfidenzintervall	Insgesamt		darunter mit Berufswechsel 95 %-Konfidenzintervall		
	in %	in %		in %	in %			
Insgesamt (n = 100 %)	(734)	17,4	14,6	20,3	(13.279)	14,7	14,1	15,3
Übergang 1996/97	30,0	17,7	12,6	22,8	32,8	14,6	13,6	15,7
1997/98	31,1	18,0	12,6	23,3	34,0	14,3	13,3	15,3
1998/99	39,0	16,8	12,4	21,2	33,2	15,3	14,2	16,4
Alter (Geburtsjahr): =18 Jahre	11,4	14,3	6,8	21,8	8,5	13,9	11,8	15,9
19-20 Jahre	54,5	18,3	14,3	22,2	53,2	14,7	13,9	15,5
21-22 Jahre	27,2	17,0	11,8	22,2	27,7	15,2	14,1	16,4
23-24 Jahre	6,8	18,0	7,6	28,4	10,6	14,4	12,6	16,3
Geschlecht: Männlich	56,7	18,8	14,8	22,7	51,0	16,6	15,7	17,4
Weiblich	43,3	15,7	11,7	19,7	49,0	12,8	12,0	13,7
Land der Betriebsstätte: West	77,1	17,7	14,4	20,9	80,4	14,3	13,6	15,0
Ost (inkl. Berlin)	22,9	16,7	10,9	22,4	19,6	16,5	15,1	17,9
Allgemeiner Schulabschluss: <Abitur	85,6	17,2	14,1	20,3	88,8	15,2	14,6	15,9
Abitur	14,4	18,9	11,4	26,3	11,2	10,9	9,3	12,5
Staatsangehörigkeit: Deutsche	91,8	17,2	14,3	20,1	93,3	14,3	13,6	14,9
Ausländer (nur West)	8,2	20,0	9,7	30,3	6,7	21,5	18,8	24,1
Ausbildungsberufe:								
Pflanzenbauer – Ernährungsberufe	34,9	18,0	13,0	22,9	30,7	17,7	16,5	18,8
Bauberufe	6,5	12,5	3,1	21,9	5,3	11,5	9,1	13,8
Raumausstatter – Techniker	6,9	25,5	13,7	37,3	8,7	16,1	13,9	18,2
Waren- und Dienstleistungskaufleute	20,4	20,0	13,4	26,6	20,6	16,4	15,0	17,8
Organisations – Erziehungsberufe	25,7	16,4	11,2	21,6	28,2	9,8	8,8	10,7
Allg. Dienstleistungs- und Verkehrsberufe	5,4	5,0	0,0	11,8	6,5	18,0	15,4	20,6

Quelle: Mikrozensuspanel 1996-1999; Beschäftigtenstichprobe (IABS-R01); eigene Berechnungen. (Tabelle zu Abbildung 6.6 im Textteil.)

Tabelle A.8: Ergebnisse des Logit-Modells zum Berufswechsel von Absolventen einer Berufsausbildung im dualen System der Übergänge 1996/97, 1997/98 und 1998/99 im Mikrozensuspanel und in der Beschäftigtenstichprobe

Merkmal (Alter bis Berufsgruppe zum Zeitpunkt ein Jahr vor dem Abschluss)	Koeffizienten (Standardfehler)	
	Mikrozensuspanel	Beschäftigtenstichprobe
Konstante (Y=1 Berufswechsel)	-1,6861 * (0,3727)	-1,7018 * (0,1001)
Übergang 1997/98	-0,0155 (0,2496)	-0,0415 (0,0609)
Übergang 1998/99	-0,0700 (0,2412)	0,0536 (0,0604)
Alter=19-20 Jahre	0,2847 (0,3431)	0,0437 (0,0939)
Alter=21-22 Jahre	0,1294 (0,3896)	0,2274 * (0,1003)
Alter=23-24 Jahre	0,0955 (0,5288)	0,2596 * (0,1198)
Geschlecht=weiblich	-0,2005 (0,2575)	-0,1736 * (0,0655)
Land/Betriebsstätte=Ost	-0,0476 (0,2449)	0,2235 * (0,0621)
Allgem. Bildungsabschluss=Abitur	0,1514 (0,3324)	-0,4697 * (0,0979)
Staatsangeh.(West)=Ausländer	0,2219 (0,3525)	0,4778 * (0,0877)
Job/Berufsgruppe 1996=Bauberufe	-0,4331 (0,4701)	-0,5382 * (0,1271)
Job=Raumausstatter-Techniker	0,4344 (0,3632)	-0,0757 (0,0907)
Job=Waren-/Dienstleistungskaufleute	0,2160 (0,2947)	0,0899 (0,0777)
Job=Organisations-/Erziehungsberufe	0,0157 (0,3187)	-0,5223 * (0,0843)
Job=Verkehrs-/Allg. Dienstleistungsberufe	-1,3285 (0,7630)	0,1401 (0,1086)
Stichprobengröße (n)	734	13.279
Log-Likelihood	-333,84	-5.450,35
Likelihood-Ratio-Statistik	G <sup>2</sup>	203,99
	d.f.	14
	P	0,00

\* Koeffizient signifikant mit  $P < 0,05$ . Wald-Test auf Gleichheit der jeweiligen Koeffizienten beider Datensätze insgesamt (d.f.=15):  $P = 0,0486$ . Bei Wald-Tests einzelner Koeffizienten (ohne Korrektur für Mehrfachtests): alle P-Werte  $> 0,05$ .



# Tabellenverzeichnis

Tabelle 3.1:	Ausbildungsstatus im April 1997 nach Beschäftigungsdauer für Auszubildende im April 1996 im Mikrozensuspanel 1996-1999 – Fallzahlen und Zeilenprozentwerte . . . . .	30
Tabelle 3.2:	Unter der Annahme vollständig zufälliger Ausfälle (MCAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte . . . . .	32
Tabelle 3.3:	Unter der Annahme bedingt zufälliger Ausfälle (MAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte . . . . .	33
Tabelle 3.4:	Unter der Annahme nicht ignorierbarer Ausfälle (NINR) geschätzter Ausbildungsstatus – Zeilenprozentwerte . . . . .	35
Tabelle 3.5:	Ausbildungsstatus im April 1997 nach Beschäftigungsdauer für Auszubildende im April 1996 im Mikrozensuspanel 1996-1999 und in der Beschäftigtenstichprobe – Fallzahlen und Zeilenprozentwerte . .	37
Tabelle 3.6:	Unter der Annahme bedingt zufälliger Ausfälle (MAR) geschätzter Ausbildungsstatus – Zeilenprozentwerte . . . . .	39
Tabelle 3.7:	Unter der Annahme nicht ignorierbarer Ausfälle (NINR) geschätzter Ausbildungsstatus – Zeilenprozentwerte . . . . .	40
Tabelle 3.8:	Unter der Annahme bedingt zufälliger (MAR) und nicht ignorierbarer Ausfälle (NINR)geschätzter Ausbildungsstatus – Zeilenprozentwerte .	42
Tabelle 4.1:	Fallauswahl beim Mikrozensuspanel für die Analysen zum Auszug aus dem Elternhaus . . . . .	49
Tabelle 4.2:	Veränderungen der Risikopopulation der 15- bis 26-jährigen Personen im Jahre 1996 durch Weg- und Zuzüge zwischen 1996 und 1997 . . . .	50
Tabelle 4.3:	Verbleib 15- bis 26-jähriger Jugendlicher des Jahres 1996 im Elternhaus nach Lebensalter und Erhebungszeitpunkt . . . . .	56
Tabelle 4.4:	Ergebnisse logistischer Regressionen zum Auszug aus dem Elternhaus .	59
Tabelle 5.1:	Besucher der gymnasialen Oberstufe 1996-1999 . . . . .	75
Tabelle 5.2:	Bildungsverlauf der Besucher der gymnasialen Oberstufe in Bundesländern mit 13-jährigem Schulbesuch insgesamt und für die 1996 16- bis 19-Jährigen (Spaltenprozentwerte) . . . . .	81
Tabelle 5.3:	Bildungsverlauf der Besucher der gymnasialen Oberstufe in Bundesländern mit zwölfjährigem Schulbesuch insgesamt und für die 1996 16- bis 19-Jährigen (Spaltenprozentwerte) . . . . .	82
Tabelle 5.4:	Absolventen allgemein bildender und beruflicher Schulen mit Fachhochschulreife und Hochschulreife 1996 bis 1998 in der Bildungsstatistik und im Mikrozensuspanel (in 1.000) . . . . .	86

Tabelle 5.5:	Antwortvariabilitäten zum allgemeinen Schulabschluss im Mikrozensuspanel 1996-1999 (Übergangswahrscheinlichkeiten zwischen den Zeitpunkten t und t+1) – Zeilenprozentwerte . . . . .	93
Tabelle 5.6:	Bis 1999 erreichter Bildungsstatus der 1996 16- bis 19-jährigen Besucher der gymnasialen Oberstufe (Zeilenprozentwerte) . . . . .	97
Tabelle 5.7:	Log-lineare Modelle zum Ausfall und Status bis 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 . . . . .	100
Tabelle 5.8:	Log-lineare Modelle zum Ausfall und Status 1997 bis 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 . . . . .	102
Tabelle 5.9:	Geschätzter Bildungsstatus im Jahre 1999 verschiedener Modelle nach der Stellung im Beruf des Familienvorstands für die Ausfallgruppe – Spaltenprozentwerte . . . . .	104
Tabelle 6.1:	Merkmalsverteilungen in der Beschäftigtenstichprobe und im Mikrozensuspanel; 15- bis 24-jährige Auszubildende in anerkannten Ausbildungsberufen im April 1996 – in Prozent . . . . .	117
Tabelle 6.2:	Übergänge 1996/97 der 15- bis 24-jährigen Auszubildenden in anerkannten Ausbildungsberufen des Jahres 1996 . . . . .	121
Tabelle 6.3:	Perzentile und arithmetische Mittelwerte der Beschäftigungsdauer im Betrieb (Stand: April 1996; in Monaten) für Auszubildende des dualen Systems im April 1996 mit Ausbildungsabschluss bis 1999 im Mikrozensuspanel und in der Beschäftigtenstichprobe . . . . .	128
Tabelle 6.4:	Log-lineare Modelle zum Status (Y) und Ausfall (R) 1997 der bis 24-jährigen Auszubildenden des Jahres 1996 (Erstausbildungen) . . . .	130
Tabelle 6.5:	Statusverteilungen im Jahre 1997 der räumlich immobilen Auszubildenden im Mikrozensuspanel des Jahres 1996 bei verschiedenen Korrekturgewichtungen . . . . .	136
Tabelle 6.6:	Abgrenzung der Stichproben für die Analysen zum Übergang von Ausbildungsabsolventen in eine anschließende Beschäftigung. . . . .	141
Tabelle 6.7:	Anteile von Berufswechseln bei Ausbildungsabsolventen der Übergänge 1996/97, 1997/98 und 1998/99 in Abhängigkeit von verschiedenen differenzierten Berufsangaben (in Prozent). . . . .	142
Tabelle A.1:	Verteilung der unabhängigen Variablen zur Analyse „Auszug aus dem Elternhaus“ in West- und Ostdeutschland (Spaltenprozentwerte) . . . .	181
Tabelle A.2:	Regressionskoeffizienten (b) und Standardfehler (in Klammern) sowie Effektkoeffizienten (exp(b)) des diskreten Hazardraten-Regressionsmodells 21 in Tabelle 4.4. . . . .	183
Tabelle A.3:	Regressionskoeffizienten und Standardfehler der log-linearen Modelle in Tabelle 5.8 zum Ausfall und Status von 1997 bis 1998 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 (Logit-Spezifikation). . . . .	186

Tabelle A.4:	Status der 1996 15- bis 24-jährigen Auszubildenden im April 1997 in der Beschäftigtenstichprobe und im Mikrozensuspanel – in Prozent . . .	191
Tabelle A.5:	Regressionskoeffizienten und Standardfehler ausgewählter log-linearer Modelle zum Status und Ausfall 1996/97 der bis 24-jährigen Auszubildenden des Jahres 1996 . . . . .	193
Tabelle A.6:	Regressionskoeffizienten gewichteter Logit-Modelle zum Statusübergang 1996/97 der bis 24-jährigen Auszubildenden des Jahres 1996 . . . . .	196
Tabelle A.7:	Berufswechsel von Absolventen einer Berufsausbildung im dualen System der Übergänge 1996/97, 1997/98 und 1998/99 im Mikrozensuspanel und in der Beschäftigtenstichprobe nach verschiedenen Merkmalen . . . . .	198
Tabelle A.8:	Ergebnisse des Logit-Modells zum Berufswechsel von Absolventen einer Berufsausbildung im dualen System der Übergänge 1996/97, 1997/98 und 1998/99 im Mikrozensuspanel und in der Beschäftigtenstichprobe . . . . .	199



# Abbildungsverzeichnis

Abbildung 2.1:	Rotation im Mikrozensus . . . . .	19
Abbildung 2.2:	Beobachtungsdauer, Zeitraum und Zusammensetzung der Stichprobe . .	20
Abbildung 2.3:	Systematik der Zusammenführung zweier Querschnitte . . . . .	20
Abbildung 4.1:	Anteil der Personen, die 1996 bei ihren Eltern leben . . . . .	48
Abbildung 4.2:	Hazardraten zum Auszug aus dem Elternhaus zu einem Modell ohne Kovariaten und für die Referenzkategorie (Modell 21) . . . . .	61
Abbildung 4.3:	Hazardraten zum Auszug aus dem Elternhaus nach Geschlecht . . . . .	62
Abbildung 4.4:	Hazardraten zum Auszug aus dem Elternhaus nach Gemeindetyp . . . .	62
Abbildung 4.5.:	Hazardraten zum Auszug aus dem Elternhaus nach dem Schulbesuch (1996). . . . .	63
Abbildung 4.6:	Hazardraten nach dem Ende des Schulbesuchs . . . . .	64
Abbildung 4.7:	Hazardraten zum Auszug aus dem Elternhaus nach Schulartwechseln (1996/97; 1996-97/98) . . . . .	64
Abbildung 4.8:	Hazardraten zum Auszug aus dem Elternhaus nach der Stellung im Beruf (1996) . . . . .	65
Abbildung 4.9:	Hazardraten zum Auszug aus dem Elternhaus nach Betriebswechseln . .	65
Abbildung 4.10:	Hazardraten zum Auszug aus dem Elternhaus nach Äquivalenzeinkommen 1996 . . . . .	66
Abbildung 4.11:	Hazardraten zum Auszug aus dem Elternhaus nach Familientyp 1996 . .	67
Abbildung 4.12:	Hazardraten zum Auszug aus dem Elternhaus nach der Geschwisterzahl 1996 . . . . .	67
Abbildung 4.13:	Hazardraten zum Auszug aus dem Elternhaus nach der Stellung im Beruf der Familienbezugsperson 1996. . . . .	68
Abbildung 5.1:	Grundstruktur des Bildungswesens in der Bundesrepublik Deutschland.	73
Abbildung 5.2:	Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 – Bundesländer mit 13-jährigem Schulbesuch (Insgesamt) . . . . .	78
Abbildung 5.3:	Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996, räumlich Immobiler – Bundesländer mit 13-jährigem Schulbesuch . . . . .	78
Abbildung 5.4:	Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996 – Bundesländer mit zwölfjährigem Schulbesuch (Insgesamt). . . .	79
Abbildung 5.5:	Status 1999 der 16- bis 19-jährigen Besucher der gymnasialen Oberstufe 1996, räumlich Immobiler – Bundesländer mit zwölfjährigem Schulbesuch . . . . .	80



Abbildung 5.6:	Besucher der gymnasialen Oberstufe im Mikrozensus 1996 und in der Bildungsstatistik (Schuljahr 1995/96) (in 1.000). . . . .	88
Abbildung 5.7:	Besucher der gymnasialen Oberstufe im Mikrozensus 1991 und in der Bildungsstatistik (Schuljahr 1990/91) – Westdeutschland (in 1.000) . . .	90
Abbildung 5.8:	Schüler allgemein bildender Schulen und beruflicher Gymnasien mit dem Bildungsziel Abitur und Fachhochschulreife im Mikrozensus 1989 und in der Bildungsstatistik (Schuljahr 1988/89) (in 1.000) . . . . .	91
Abbildung 5.9:	Bis 1999 erreichter Bildungsstatus der Besucher der gymnasialen Oberstufe des Jahres 1996 nach beruflicher Stellung des Familienvorstands und Alter . . . . .	98
Abbildung 5.10:	Absolventen allgemein bildender Schulen (bis 21 Jahren) mit Allgemeiner Hochschulreife 1996 nach Region und Alter (Geburtsjahr) (in 1.000). . . . .	106
Abbildung 5.11:	Geschätzter Bildungsstatus im Jahre 1999 für die 1996 16- bis 19-jährigen Besucher der gymnasialen Oberstufe einschließlich Ausfällen (in Prozent) . . . . .	107
Abbildung 6.1:	Kurzbeschreibung zur Abgrenzung der Auszubildenden im dualen System und der Ausbildungsabschlüsse . . . . .	115
Abbildung 6.2:	Status der 1996 15- bis 24-jährigen Auszubildenden im April 1997 in der Beschäftigtenstichprobe (B) und im Mikrozensuspanel (M) - in Prozent. . . . .	124
Abbildung 6.3:	Modellergebnisse zum 1997 geschätzten Status (Y) der bis 24-jährigen Auszubildenden des Jahres 1996 - in Prozent. . . . .	133
Abbildung 6.4:	Geschätzter Ausbildungsstatus im April 1997 (Y) der bis 24-jährigen Auszubildenden des Jahres 1996 mit einer mehr als zweijährigen Beschäftigungsdauer - in Prozent. . . . .	134
Abbildung 6.5:	Anteile des Status 1997 in der Beschäftigtenstichprobe (B) insgesamt und für räumlich immobile Auszubildende des Jahres 1996 im Mikrozensuspanel (M) für ausgewählte Merkmale bei verschiedenen Modellannahmen – in Prozent. . . . .	137
Abbildung 6.6:	Berufswechsel von Ausbildungsabsolventen im Mikrozensuspanel und in der Beschäftigtenstichprobe (in Prozent). . . . .	145



Mit der Zusammenführung der Querschnittserhebungen des Mikrozensus zu einem Paneldatensatz eröffnen sich für die Forschung neue Potenziale. Es entstehen jedoch Probleme durch Panelausfälle, da die aus dem Auswahlbezirk wegziehenden Haushalte und Personen nicht weiter befragt werden. Diese Arbeit konzentriert sich auf Bildungsverläufe von Jugendlichen, die überdurchschnittlich räumlich mobil sind. Die Selektivitätsfragen werden am Beispiel von drei Übergängen untersucht: Dem Auszug aus dem Elternhaus, dem Abschluss der gymnasialen Oberstufe und der beruflichen Ausbildung im dualen System. Zur Klärung selektiver Ausfälle werden Referenzdaten der amtlichen Bildungsstatistik und die Beschäftigtenstichprobe des IAB sowie Selektionsmodelle verwendet. Die Analysen zeigen, dass Panelausfälle bei Verlaufsanalysen zum Besuch der gymnasialen Oberstufe mithilfe von Gewichtungsfaktoren weitgehend korrigierbar sind. Für berufliche Ausbildungsverläufe trifft dies nicht zu. Ergänzend werden Analysen zur Antwortkonsistenz der Angaben zum allgemeinen Schulabschluss zu Fragen der Datenqualität durchgeführt.

Several cross-sections from the German Microcensus have been merged generating a rotating panel dataset which will provide the scientific community with a rich source of longitudinal data. However, there are problems with panel attrition due to households that moved out of the sampled areas not being tracked. The primary objective of this research is to investigate educational transitions made by highly mobile young people. Selective panel attrition is examined for three transitions: leaving the parental home, completing the second stage of academic secondary education, and completing an apprenticeship. The analyses use data from official education statistics and the IAB Employment Sample as reference statistics, and selection models are estimated. For transitions from academic secondary education the selection bias can be corrected by means of weight adjustments. However, this does not apply to transitions of apprentices from the dual system of vocational training. In addition, analyses on item stability of data on general educational certificates provide information on the data quality.

**gesis**

Lennéstraße 30 • 53113 Bonn  
Telefon: 02 28 / 22 81 - 0  
Telefax: 02 28 / 22 81 - 120  
[info@gesis.org](mailto:info@gesis.org)

Ein Institut der Leibniz-Gemeinschaft

ISBN 978-3-86819-003-8  
Schutzgebühr: EUR 15,-